Agents, affectivity and aesthetics in user-interface design

Karen Wilson

A thesis submitted in fulfilment of the requirements of the University of Abertay Dundee for the degree of Doctor of Philosophy

June 2007

I certify that this thesis is the true and accurate version of the thesis approved by the examiners.

Signed [redacted] ..... **Date**..20/06/07

Acknowledgements

I would like to thank Leona Elder for her supervision, help, support, and encouragement throughout my thesis, especially in the last year when she helped me turn this into a proper story and made me believe that I could do this. Without her, I would not be writing these acknowledgements because there would be no thesis. For this, I am forever grateful.

Secondly, I would like to thank Lynn Wright for keeping me sane throughout this, and for her help in making sure everything made sense, triple-checking my references, and helping to piece everything together. I would like to thank her for her continued friendship too (nearly 28 years Sate!! ☺) and for letting me disappear and read crime when I'd had enough of the PhD!

I am indebted to these two people and could not have done this without them.

Thanks to Malcolm Cook for his supervision and support throughout my thesis. He has changed me in a lot of ways, and I am grateful for that!

Thanks also to Scott Hardie for always reassuring me that I could get this finished, and providing copious amounts of gin when necessary!

I would also like to thank my parents for supporting me throughout this. I know they thought it was yet another excuse not to get a 'proper' job... and they were right! Without their encouragement I would not have got through this, and I owe them a huge amount of thanks.

I would like to thank NCR, Dundee, for funding the first 2 years of my studies, and in particular Lynne Coventry and Graham Johnson for their support.

Thanks to everyone in the Epicentre for making it such a nice place to work. Special thanks to Colin Cartwright for employing me for the past few years, which allowed me to continue with my studies, and Tudor Radulescu for writing my agent software.

I would also like to thank Dr Ken Scott-Brown and Dr Sandra Cairncross for making my viva an altogether pleasant experience. Their positive comments have been a huge confidence boost!

Finally, thanks to Sate, Whale and Bag... 2007 will be a better year! ☺

Contents

List of Figures

List of Tables

List of Appendices

Abstract

Human-Computer Interaction (HCI) research has suggested that there has been a general shift away from more traditional aspects of design, such as usability, and the focus is now more pleasure, or emotion, based. Jordan (2000) states that pleasure-based design is associated with more hedonic aspects, such as enjoyment, and the interaction between the user and the product. A number of studies have investigated the relationship between the user and product in relation to agents (e.g. DeAngeli, Lynch & Johnson, 2002). While it has been widely acknowledged that the use of agents enhances the user experience (e.g. Lester, Converse, Stone, Kahler & Barlow, 1997; DeAngeli et al, 2002), these agents have mainly been fully animated, which may influence perceptions as they are considered more 'human-like'. Conversely, other studies (e.g. Koda & Maes, 1996) have found that the appearance of an agent has little influence on user perceptions of attributes such as intelligence, but only if an interaction has taken place. A series of experiments were conducted in order to investigate user perceptions of agents, or more specifically the influence of aesthetics, context and interaction. Experimental work investigated rating of agents, on a variety of different attributes, in an implied financial context. It was found that there was a general positive regard for female agents, but it was unclear whether stereotypes relating to the context were driving these judgements. Additional work showed that this positive view of female agents was consistent when no context was implied. Further investigation of the role of context indicated that while context (compared to imagined context and no context) had an overall detrimental influence on perceptions of agents, there was a high general regard for both attractive agents and female agents. However, to ensure that this result did not simply arise from the choice of website (financial), additional studies investigated the extent of occupational stereotypes and whether these occupational stereotypes extended to agents. This was found to be the case. However, although the most appropriate agent for a given occupation was one that was gender-congruent with the occupation, there was still a general positive regard for attractive agents. Finally, the influence of interaction was explored, and it was found that the aesthetics, or even presence, of an agent had no effect on user perceptions. This suggests that the quality of interaction may be the most salient aspect of agent perception. These findings are discussed in relation to the literature and future studies are considered.

Chapter 1 : Beyond Usability – Pleasure and Emotion in Design

1.0. Aims of the thesis

One of the most salient factors when judging other people is attractiveness. Many studies investigating attractiveness have found that the "what is beautiful is good" stereotype to be a common phenomenon. HCI literature suggests that similar patterns are emerging with regard to interface design, with the focus shifting away from traditional usability aspects and towards the aesthetic and pleasurable/emotional aspects of design. The use of agents in HCI has been received both positively and negatively, with many of the inconsistencies among studies being attributed to the type of agent being used and the task being carried out. However, very few studies have reported differences in perceptions of agents based upon their physical appearance. Therefore, this thesis will investigate the influence of aesthetics on perceptions of agents. Possible underlying influences such as context, occupational stereotypes, and interaction will be investigated.

1.1 Introduction

Traditionally, system, or product, acceptance was judged to be based mainly on aspects of usability. In laypersons' terms, usability means ensuring that a product is easy to use. More formally, the International Standards Organisation (ISO) defines usability as "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments" (ISO DIS 9241-11), where 'effectiveness' is defined as the extent to which a goal, or task, is achieved; 'efficiency' is defined as the amount of effort required to accomplish a goal; and 'satisfaction' is defined as the level of comfort that the user feels when using a product, and how acceptable the product is to users as a means of achieving their goal.

However, Nielsen (1993) believes that usability, together with utility, influences the usefulness of the product and usefulness is one of the attributes affecting acceptability. Similarly, Shackel (1991) considers usability, along with utility, likeability, and cost to be influencing factors in product acceptance. Thus, product acceptance may be based upon many different factors, with usability itself not being

as straightforward as simply being "easy to use". More specifically, if the dimensions of usability are considered, user satisfaction, for example, may be particularly difficult to determine simply by asking if a product is easy to use. Therefore, in order to fully understand what makes a product acceptable to users, other aspects of product design must be considered.

1.2 Pleasure-based design

Although there has been extensive research carried out concerning aspects of product usability (e.g. Nielsen (1993), Scholtz, Laskowski, and Downey (1998), and Shackel (1991)), other aspects such as aesthetics, sensitivity, and pleasure are beginning to be considered important for product use (Jordan, 1998). Until recently, traditional human factors thinking has concentrated predominantly on performance related issues in order to understand how people interact with products and systems. There is no doubt that usability is an important issue, but the main focus of a usability approach concerns the product as merely a tool with which to complete a task. However, there has been a shift recently to look at the enjoyment, or emotional, aspects of product and system use. Issues such as appeal, pleasure, fun, and taste are all being considered in the context of product design. Jordan (1997a) suggests that products mean more to people than just something with which to complete a task; they are seen as living objects with distinct personalities, with which people form relationships. He suggests that products can therefore engender a whole host of positive and negative emotions such as happiness, security, or frustration. Jordan also suggests that designers need to aim higher than 'satisfaction' in product design, and set their sights on 'pleasure'. He states that the emotional and hedonic issues should be considered "in combination with the practical aspects, not at their expense" (p. 249).

The area of 'pleasure-based human factors' is expanding rapidly. In addition to the growing body of literature addressing issues related to pleasure with products, there is an increasing number of conferences dedicated to human factors, and more specifically to the area of 'pleasurable products'. Jordan (1997b) defines pleasure in product use as being "the emotional and hedonic benefits associated with product use" (p. 35). Some of the key issues in what makes a design pleasurable to use include the relationship between the user and the object, and the dichotomy between liking the

familiar and wanting surprise and innovation: products cannot be isolated from their context and the previous experience of users, and some of the emotionally appealing aspects of a product or interface may reduce usability. Jordan (1998) found for a number of products that those experienced as pleasurable were used more than they would be otherwise. Similarly, products which were not pleasurable were used less because of the displeasure associated with the product. He concludes that, although usability is important to how pleasurable a product is, "the issue of pleasure in product use also goes significantly beyond usability" (p.32).

It is relatively simple to determine whether or not a product is usable[1], but it is far more difficult to identify what makes a design or product pleasurable or enjoyable to use. As technology is now so advanced, manufacturers are producing high quality, usable products, and product design is one of the few ways in which manufacturers can gain an edge over their competitors. Users now generally have the expectation that products will be easy to use. Consequently, pleasure and the emotional aspects of design are becoming increasingly important factors in product design. For example, in their top fifteen reasons "why you'll love a Mac", Apple list at number eight "design that turn heads", which includes not only the software and desktop, but also the design of the hardware itself (Evans, 2007). Similarly, Oh and Khong (2003) stated that customers expect quality and functionality in their products and now demand more 'pleasurable' features in the design. This, they say, is one way a company can gain competitive advantage. However, they found that that introduction of pleasure in the workplace was more beneficial in competitive terms that introducing pleasure in the product itself. Pleasure in the workplace, they claim, leads to better performance, which in turn leads to superior quality products.

Overbeeke, Djajadiningrat, Hummels, and Wensven (2002) proposed that a product's functions should be aesthetically pleasing, and that interacting with the product should contribute to overall pleasurability. They go on to suggest that there is more to usability than being easy to use. Many people may choose to use a product even if it is difficult to use because they find the experience stimulating, challenging, rewarding, and therefore enjoyable (for example, learning to play a musical

[1] Usable = Alpha Usability. i.e. it can be used to achieve the goal

instrument). Snelders (1995) (cited from Creusen and Snelders, 2002) showed that pleasurable and hedonic product values and rational involvement with products are two separate, unrelated aspects of the general involvement with products. That is, pleasure is a separate but equally important aspect of product involvement as are the actual product functions. Creusen and Snelders (2002) found that pleasure in products is often based on form or appearance, but that it is difficult to design a product to convey a specific impression as "pleasure-based human factors are based mostly on holistic impressions and ... these are communicated mostly by abstract attributes" (p.73). For example, the Apple iPod is the biggest selling MP3 player in the UK and US (Evans, 2007), with one of its key features being the somewhat abstract shuffle wheel.

In a similar line of research, Russo and de Moraes (2003) suggest that many designers have now started focussing more on the aesthetic aspects of a product, at the expense of traditional ergonomic and usability tests. One such example of this, they suggest, is Philippe Starck's *Juicy Salif*; a product that many people purchase as a "sculpture without the pretension of being an art object" (p. 146). This became even more apparent with the launch of the anniversary gold edition, which couldn't actually be used as the citric acid in the lemon juice would cause it to rust. Russo and de Moraes carried out a simple usability tests using the Juicy Salif and found that while most people like the look of the product they would not trade it for their existing lemon juicer, with 4 out of the 6 people tested stating that it was difficult to use.

1.3 Product Appeal

Hassenzahl, Platz, Burmester and Lehner (2000) suggest that while traditional usability aspects provide an effective and efficient way to design software systems, something else is needed to make it novel and surprising. They suggest that the combination of a system being usable and interesting could increase the appeal of the system, thus providing the user with a greater sense of enjoyment. In their study, they found judgements of appeal before and after using a system to be highly consistent. Regression analysis revealed that ergonomic qualities, such as efficiency and effectiveness (i.e. traditional usability factors), and hedonic qualities such as novelty, innovativeness, and aesthetics, both contributed equally to judgements of appeal

before and after use. However, further analysis revealed that there was a significant interaction between usage time (pre and post use) and quality type, with hedonic qualities significantly increasing after use and ergonomic qualities decreasing after use. Judgements of appeal remained constant before and after use. Hassenzahl and colleagues suggest two possible implications of this. Firstly, it shows that both qualities can be influenced by interaction, and secondly an increase in one quality may always result in a decrease in the other quality. However, judgements of appeal remained consistent, thus suggesting that both qualities must be taken into account when designing a software system. But, if a system is lacking in ergonomic qualities this can be compensated by an increase in hedonic qualities, and vice versa.

1.4 Emotional Design

It is widely acknowledged that humans are not entirely rational and are strongly influenced by their emotions (Goleman, 1996); thus, Jordan (2000) suggests that human emotion should play an important element in product design. Norman (2004) asserts that emotion has to be taken into account in product design, simply because it (emotion) is such an important part of people's everyday lives. He states "utility and usability are important, but without fun and pleasure, joy and excitement, and yes, anxiety and anger, fear and rage, our lives would be incomplete" (p. 8). Not only are emotions critical to human behaviour, Norman suggests that they are of as equal importance for intelligent machines, particularly autonomous machines.

It has been shown that positive emotions such as happiness can actually expand people's thought processes and facilitate creative thinking (Isen, 2000). This, Norman proposes can make products seem easier to use; aesthetically appealing products make people feel good, and feeling good makes people think more creatively and therefore become better at brainstorming, solving problems, and generally exploring various alternatives to problems; thus, products will be easier to use as it will be easier for people to find alternatives if their first attempt at using a product fails. This also means that, because a person or user is more creative, they will be able to overlook or deal with minor problems when using a product or system, particularly if they find it fun or pleasurable to work with. Conversely, when people are anxious or angry, they focus more on minor details and faults with a product or system. Thus, designers

must ensure information on how to use the product is easily accessible, with clear help or feedback if required.

Design preference can also be influenced by factors such as the occasion in which a product will be used, the context in which it will be used, and the mood of the person using it (Norman, 2004). Norman suggests that products are much more than just utilitarian objects; they are art, often having personal meaning to people.

Norman proposes that there are several components to product design: usability, aesthetics, and practicality. In addition to the practical aspects involved in creating a product, such as material, the way in which it is manufactured, marketing method, cost and practicalities, and how usable the product is, Norman suggests that the "strong emotional component to how products are designed and put to use" may be more crucial to creating a successful product than the practical elements.

1.5 Interaction

Jordan (2000), however, contends that pleasure from products comes not from the product itself, but from interacting with the product, and the relationship formed between a person and a product. Jordan proposes a four-scale framework for designing pleasurable products, based on Tiger's (1992) definition of pleasure: psycho-pleasure, physio-pleasure, socio-pleasure, and ideo-pleasure. These four pleasures encompass the interactions and feelings a person has when using a product, where Psycho-pleasure deals with pleasures of the body such as sights, sounds, smells, tastes, and touch; Physio-pleasure is derived from people's reactions and psychological state during the use of products; Socio-pleasure deals with interaction with others, with many products playing important social roles, either by design or by accident; and Ideo-pleasure reflects the experience a user has with a product, appreciating the aesthetics, or the quality, or the extent to which a product can enhance the user's life. In essence, pleasure based design means that all the potential benefits must be taken into consideration when designing a product.

Similarly, Hummels (1999) suggests that in order for a product to be engaging, the focus in the design of the product must not be solely on the product but rather on the relationship between the user and the product. That is, a product should not only be

designed to have a beautiful and pleasing appearance, it should be designed so that the user's interaction with the product is beautiful and pleasing.

The process of designing pleasurable products requires designers to take into consideration all the possible interactions a user might have with a product, carefully balancing the effects of functionality, usability, and aesthetics. Once there is an appropriate balance of these three factors, users will be more likely to interact with them on a regular basis, form relationships with them, and tell other people about them (Forlizzi & Ford, 2002).

1.6 Aesthetics

Aesthetics is one quality that can also contributes to a user's overall experience when interacting with a product (Alben, 1996). Stolterman (1994) conducted a study into how aesthetics relate to information systems, and suggested that the experience of using and interacting with a system cannot be explained simply in terms of functionality and usability, but that other factors such as aesthetics can influence a user's overall perception of the system.

Interacting with an aesthetically pleasing product ensures emotional and sensory experiences, as opposed to usable products that address the users' mental model. A product that has aesthetic appeal will often become meaningful and idiosyncratic to the user (Forlizzi, Hirsch, Hyder & Goetz, 2001).

Norman (2004) suggests that there are different levels of design, each of which relates to different aspects. Visceral level design is concerned with appearances, people's initial reaction to a design and the immediate emotional impact, and the enjoyment a person gets from a product because of its appearance. Many people purchase products based on looks alone, with many similar products being turned down because they are not as aesthetically appealing. Behavioural level design relates to the pleasure and effectiveness of use; the performance of the product. Finally, Reflective level design is about the 'meaning' of a product and/or its use, and about the message a product sends to others.

While attractiveness is a visceral-level experience, beauty is a reflective-level phenomenon. Beauty is more than just how an object or product looks on the surface;

it is shaped by experience, knowledge, and culture. An example of this difference can be seen in advertising, which can appeal at the visceral or reflective level. Attractive products, such as cars, Apple iMacs, and "sexy" bottles for drinks and perfume, all work at the visceral level. Reflective level advertising plays more on the exclusive, prestige, and perceived rarity of products, such as expensive alcohol, or the exclusiveness of a club or restaurant. Another major factor at the reflective level is customer relationships. If a company builds up a good relationship with its customers, offering a high level of assistance, this can, in effect, reverse any negative experiences they may encounter with a product. The whole "customer experience" is at the heart of reflective design, providing that extra special touch and a warm, friendly interaction. These experiences can influence further purchases or usage, as well as recommending the product to others (Norman, 2004).

1.7 Aesthetics and Perceived Usability in User Interfaces

Kurosu and Kashimura (1995) studied ATM layout patterns in terms of perceived functionality and aesthetics in Japan. Twenty-six screens were evaluated by participants on how easy they looked to use and how beautiful they were. The results showed a significant positive correlation between beauty and apparent usability. Professional interface designers were then consulted in order to discover which factors they considered enhanced inherent usability. The results showed that inherent usability (as defined by the interface designers) was less correlated with apparent usability than apparent beauty was, which suggests that the aesthetics of an interface are an important factor to users even when they are evaluating the functional aspects of it.

Tractinsky (1997) replicated the study by Kurosu and Kashimura using the same ATM layouts[2], in order to test that the results were consistent in a different culture. Results showed a high positive correlation between apparent usability and aesthetics, and Tractinsky concluded that aesthetics are closely related to apparent usability, which in turn suggests that aesthetics may affect system acceptability. These results

[2] These were translated into Hebrew where necessary

also confirm that the link between aesthetics and usability is sustainable across different cultures.

Tractinsky, Katz and Ikar (2000) further investigated the relationship between perceived usability and aesthetics, both pre- and post-use. Pre-experimental measures indicated a strong positive correlation between the system's aesthetics and perceived usability. This suggests first impressions are important, in that the more attractive an interface is then the more it is perceived to be easy to use. The amount of information presented on the screen was weakly, and negatively, correlated with aesthetics and apparent usability. Post-experimental measures indicated that the strong positive correlation between aesthetics and apparent usability remained intact. Post-experimental satisfaction was highly correlated with both apparent usability and aesthetics.

The results demonstrated that users could differentiate between various properties of the system. For example, users did not associate the amount of information on the interface with the interface's aesthetics or usability. It was also found that post-use perceptions of usability were not affected by the actual usability of the system, but by the interface's aesthetics, thus suggesting that 'beautiful is usable' and that perceptions of the system's aesthetics affect perceptions of the interaction with the system.

In a study investigating how to make assistive technologies and products more pleasurable for elders, Forlizzi and colleagues (2001) state that two essential characteristics of assistive products are aesthetics and their ability to deliver pleasure. They go on to assert that not only are aesthetics important in making a product visually pleasing, they also offer a further dimension to usability. "Traditional usability factors determine whether a device *can* be used; aesthetic factors determine whether a device *will* be used, and what the emotional, psychological, and social outcomes of the user-product interaction will be" (p2).

1.8 Aesthetics and Websites

In light of the ever-changing web environment, Hudson (2000) suggests that there needs to be an integration of usability and aesthetics. He states that even if a website is highly usable it may still fail if it is not fulfilling the greater needs of the user. A website must also promote a sense of trust and security, as well as being entertaining or providing a high level of service. He recommends that websites should be designed to be elegant as well as usable.

Schenkman and Jonsson (2000) investigated how users aesthetically experienced web pages, which kind of web pages users preferred, and what factors determined the overall impression of a web page. They believed that the first impression a home page gave would be important for the appeal of, and attitude towards, that company, as well as determining whether or not the user would stay and browse the site. Participants were firstly shown pairs of websites, and for each pair asked to indicate their judgement of similarity and preference between the two web pages. The results of the similarity and preference judgements revealed that, although participants judged web pages to be similar based on the amount of complexity, legibility and order, their preferences were based on beauty. In the second task participants were asked to judge each website individually on seven bipolar category scales (these were complexity, legibility, order, beauty, meaningfulness, comprehension, and overall impression). Multiple regression analysis revealed that the best predictor of the overall impression of a website was beauty. Further analysis revealed two main category groupings: the first contained complexity, legibility and order, whereas the second contained the more semantic-related variables of meaningfulness, beauty and overall impression, indicating the importance of beauty and meaningfulness for the overall impression. The main findings in the results indicated that the beauty of a web page is an important factor determining how it will be experienced and judged.

In a similar study Tarasewich, Daniel and Griffin (2001) suggested that in addition to users finding a website easy to use, the whole experience of using a website should be enjoyable. One of the ways in which Tarasewich and colleagues propose the perception of a website and enjoyment in its use could be influenced is by employing better aesthetics. If a user is enjoying the experience of using a website, they may

browse the site for longer, return to the site at a later date, or recommend it to others. Tarasewich et al. investigated the relationship between aesthetics and usability, as well as the impact aesthetics had on the overall impression of a website. They found that while aesthetics play a part in the overall experience of using a website, other factors such as usability and context seem to be more important. However, they also found that aesthetics played a large part in making the whole experience more enjoyable, and may encourage the user to return to the site. Another way in which aesthetics are important in web design is that good aesthetics can encourage or promote interactivity, which may encourage the user to stay on the site longer, and also "promote a feeling of security and well-being that can lead to trust in the website and in the organisation" (p.14).

De Wulf, Schillewaert, Muylle and Rangarajan (2006) investigated the effect of pleasure in web page success. They found that in order for a user to experience high levels of pleasure when using a website, it must be well-organised and state-of-the-art, and have sufficient quality information. If the user experienced high levels of pleasure, then this in turn promoted a high degree of satisfaction with the website, as well as raising levels of commitment and trust. They suggest that designers should strive to create websites that induce feelings of pleasure in order to contribute to the overall satisfaction, commitment and trust with the website.

Norman (2004) states that while usable products are not necessarily pleasurable, and pleasurable products are not necessarily usable, there is no reason why usability and pleasure cannot both be present in a design. He goes on to state "we now have evidence that aesthetically pleasing objects enable you to work better ... are easier to deal with and produce more harmonious results" (p. 10).

1.9 Anthropomorphism

People frequently read emotional responses not only in almost anyone they come into contact with, but even inanimate objects which are often experienced in human terms. This ascription of human emotions, beliefs and motivations to inanimate objects or animals is known as anthropomorphism. People have a tendency to anthropomorphise the things they come into daily contact with, from their pets and

toys (such as teddy bears), to almost anything that they interact with such as cars and electrical appliances (Reeves & Nass, 1996).

It has been widely recognised that people have a relationship with the products that they use (Jordan, 1997b), which often leads to anthropomorphism, thus there is an expanding amount of literature on the personification or personality of products. Crozier (1994) suggests that the combination of use and functionality of a product may lead to affection for that object. Behavioural, sociocultural, and psychosocial influences, in addition to the task-related responses, affect the hedonic values associated with the person-product relationship. Product Personality Assignment (PPA) evaluates products by assigning human personality characteristics to products (Jordan, 1997b). Research into PPA found that participants had common perceptions about the personality of products. For example, products that were geometrically simple in design were thought of as being sensible and trustworthy, whereas more organically styled products were seen as being friendly, intuitive and cute. The colours and materials also influenced people's personality ratings. 'Extroverted' products were associated with light metals, compared to dark plastics that were associated with 'introverted' products. This research also showed that products reflecting participant's own personalities were significantly preferred over the other products in the study. Thus, Jordan suggests that it is important to understand people in a holistic manner, and, in doing so, product designers will gain a better understanding of what will appeal to their target audience.

Computers are another example of objects people often treat as human. In human-human interaction, people often rely on appearance cues/characteristics to judge what others are like, and this often influences the relationships they form. As Reeves and Nass (1996) found, people's social responses to computers mimic their social responses to other human beings. Their studies showed that not only do people respond socially to media, but that they are not aware of doing so, and believe that it is not rational to do so – many participants in their studies were not aware of reacting in a social manner. One surprising result was that these responses to media were not just happening in 'advanced' new media (e.g. virtual reality, IMAX cinemas, etc), but also to the simplest of media such as text and pictures, on standard televisions and

PCs. They also found that these unconscious social responses were made by most people, regardless of factors such as age or experience.

One of the main benefits of considering the 'Media Equation' (media equals real life, Reeves & Nass, 1996, p. 5), in relation to this thesis, is that the current research utilised similar methods to those used in human-human psychology to investigate how people interact with others. Reeves and Nass suggest that media are often thought of as tools, which can make the generation and absorption of design rules difficult. However, if people respond to media in a social, natural way, then, logically, interaction becomes more fluent and there is less demand for instruction, which in turn makes it more enjoyable and fun to use.

People often succumb to the social dynamics with computers. Norman (2004) purports that "basically, if something interacts with us, we interpret that interaction; the more responsive it is to us through its body actions, its language, its taking of turns, and its general responsiveness, the more we treat it like a social actor" p. 135.

Therefore, machines that can display emotions present a rich and satisfying interaction for the user. However, the richness and satisfaction does not actually come from the machine; it comes from the user, and their understanding and interpretation of this interaction.

1.10 Appeal of Agents

Clarke, Jordan and Cockton (1995) examined the affect of agents in an information retrieval system. The agents were used in an attempt to assess whether they made the system more engaging and pleasurable to use. This system was evaluated in comparison to a traditional information retrieval system which was hypertext based (accessing new screens by clicking on highlighted links), and both systems were evaluated in terms of usability, engagement, and overall pleasurability. The new (agent based) system was designed in accordance with the approach taken by Laurel (1993), in which she placed the six elements of a play[3], as defined by Aristotle (trans. 1954, cited in Laurel, 1993), in the context of human-computer interface design. The

[3] Action, character, thought, language, melody and spectacle.

level of engagement was measured using think aloud protocols, with the participants saying what they were doing and how they were feeling while using both systems. After using the system, participants then completed a usability questionnaire, and answered questions about their experience of using the system. Although the usability scores showed no difference between the two systems, the agent-based system was unanimously preferred and was considered to be more pleasurable to use. Further analysis of the questions revealed that the traditional system was simpler and easier to learn compared to the new agent-based system. However, participants indicated that they would be more likely to use the new system. Three agents were used in the new system; one concerned with the playwright and the play; one with the production of the play; and the final one with reaction to the play. No mention was made as to how these agents were chosen, the gender of the agents, and what they looked like. However, the study was conducted in the mid-1990's when agents were relatively novel. Since then, more applications and websites are using agents; therefore physical appearance may now be more of an influencing feature on acceptance or enjoyment.

However, the idea that using an agent in an interface makes it more engaging could be debatable. As research in social psychology has shown, if the person (or in this case the agent) was unattractive then users may not enjoy the experience of engaging with it as much and may also attribute other negative traits to it, and may find it more annoying rather than engaging in the interface. It may be possible to determine whether a certain type/design/appearance of agent engenders the sought after impression by looking at the person perception literature in social psychology. This will be discussed in further detail in Chapter 3.

1.11 User Response to Agents

In recent years, it has become apparent that due to the rate at which the Internet is growing, the impact is going to increase further in personal and business terms. Thus, Kelly (1999), and Lynch, Emmott and Johnson (1999) proposed that the focus should be on developing the relationship between the consumer and the information provider or interface. The NCR Knowledge Lab termed this 'relationship technologies' to "define existing and emerging technologies and models that enable, support and

enhance relationships between customers and providers" (p.98). In contrast to the traditional cognitive aspects of IT, relationship technologies are concerned with developing and sustaining relationships between the user and the interface. It is suggested that systems designed around 'relationship technologies' will change the traditional focus within HCI and introduce aspects such as emotions, affect, social intelligence, enjoyment, pleasure and humour in the interaction with computers. From a user's perspective, the difference between cognitive aspects of IT and social aspects is significant. De Angeli, Lynch, and Johnson (2002) suggest that in order to maintain the user/interface relationship, the agents used in the interface need to be socially responsive, empathic, vibrant, and elicit a sense of personality. In doing this, the anthropomorphic perception educed by computers can be improved.

De Angeli et al. (2002) suggested that due to the increased use of embodied agents, interactive systems, and "interfaces with personality"(p.98), the way in which users interact with computers in the future will change considerably. Thus, they suggested that the general usability framework must be re-examined in order to fully understand how people respond to these novel interfaces.

Although there has been a considerable amount of research into the concept of personality in agents, it is not yet clear what type(s) of personality are most appropriate or preferred. However, due to the complexities involved in evaluating the effect of social artifacts, De Angeli et al. proposed a framework that aims to evaluate the effectiveness of social artifacts. This framework is based on the premise that as people treat, and interact with computers in a similar way to other humans (Reeves and Nass, 1996), then tests and questionnaires from social psychology or personality research can be used as an evaluation framework for social artifacts.
The main focus of the study carried out by De Angeli et al. (2002) investigated how the introduction of a character in an interface would effect the amusement and enjoyment of an ATM (Automated Teller Machine). The character was called 'Granny' and users could interact with her, not only as a means of withdrawing cash, but also on a personal level; Granny could give account information, a personal diary, shopping list, and local maps and information.

In order to test their "evaluation framework", De Angeli et al. examined user reaction to a social artifact ('Granny') compared to a traditional ATM (cognitive artifact) and a bank cashier (human teller). They hypothesized that there would be a difference in the perception of the social artifact compared to the cognitive artifact and the human teller, but that the responses to the social artifact would be positive. Participants were asked to describe each of the three targets, and analysis suggested three distinct dimensions: functional, aesthetic, and social. The functional dimension mainly represented aspects such as how to use the target, effectiveness, and utility, i.e. it was seen as a traditional usability dimension. The social dimension represented the relationship between the user and the target, and adjectives used to describe the target related to personality traits, emotions, and attitudes. The aesthetic dimension represented the general physical appearance of the target: how it was designed and how it looked. The traditional ATM was described mainly in terms of functional qualities, whereas 'Granny' was described in mainly social terms. De Angeli et al. (2002) suggest that this difference is important in terms of the reaction elicited from the users. Social stimuli are much more difficult to measure and the reaction will change depending on the interaction, and because many of the attributes are not visible (e.g. perceived traits and attributes). 'Granny' was also judged in more social terms compared to the human teller, but De Angeli and colleagues explain this by suggesting that the judgements made about the human teller relate more to the occupation of bank cashier rather than to the actual person. This was further supported in the relatively low perceptions in social and aesthetic terms of the human teller, compared to the positive functional evaluation. The overall impression of 'Granny' was very high, particularly in social and aesthetic terms.

Based on these results, De Angeli et al. (2002) proposed an 'Involvement Framework' in order to try to evaluate the emotion and personality of social artifacts, as well as the user's relationship with them. Involvement refers to the relationship and encounter between the user and the social artifact, and the Involvement Framework is based on the three dimensions (functional, social and aesthetic) found in the above study, although the influence of each dimension will vary according to the task, context, and the user. As shown in Figure 1.1, user satisfaction is determined by the combination of each of the three dimensions.

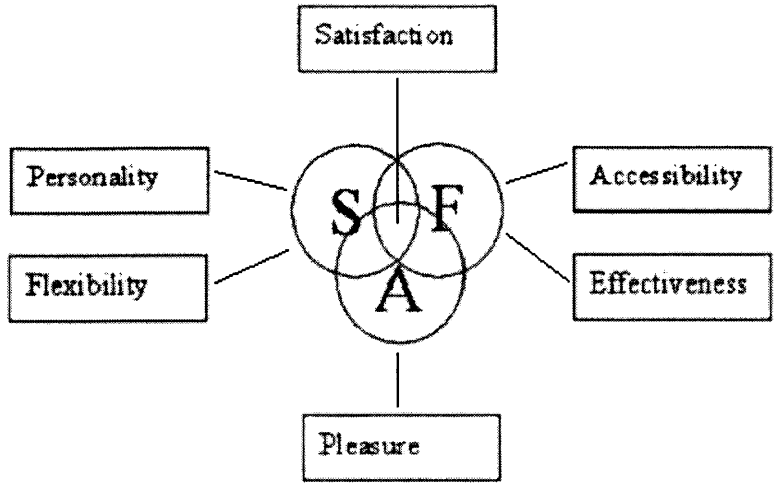


Figure 1.1: The Involvement Framework (De Angeli et al., 2002)

S=Social, F=functional, A=aesthetics

Accessibility (which refers to the effort required to communicate easily with the social artifact) and Effectiveness (which, in addition to being the main traditional usability dimension, includes factors such as amount and quality of interaction with the social artifact) encompass the functional qualities regarded important for social artifacts. The fundamental characteristic in the aesthetic dimension is pleasure, which covers all aspects of the enjoyable interaction with the social artifact. In order to enhance the user's enjoyment, the physical appearance of the social artifact should be made attractive and appealing. Two social qualities were thought to be important for social artifacts: Flexibility and Personality. Flexibility refers to the extent to which both the user and the system can adapt; that is, the user must be able to personalise the system, and the system must be able to adapt it's behaviour depending on the user. Finally, in order for a social artifact to enhance the interaction, the personality it displays must be consistent and predictable, both over time and across different contexts.

De Angeli and colleagues concluded that social artefacts, such as Granny, elicit a different set of responses from users compared to either a traditional ATM or bank cashier, and propose that the Involvement Framework is a starting point for evaluating these social artifacts.

1.12 Conclusions

As previous research has shown, visual design has a significant impact on apparent usability (Kurosu and Kashimura, 1995; Tractinsky, 1997; and Tractinsky et al., 2000). Hassenzahl, Platz, Burmester and Lehner (2000) also suggest that the user does not always perceive objective usability. For example, something designed to be simple has failed if the user perceives it as complex.

In order for a product to be both usable and pleasurable, there has to be some compromise between different aspects of the design. For instance, it is difficult for a product to be both innovative and familiar, or for a computer game to be easy to use and fun or stimulating. There is a positive correlation between user preference and exposure, but this decreases due to habituation, thus stylistic change is often cyclic (Crozier, 1994; Hassenzahl et al., 2000).

The move away from traditional usability to designing products that are a genuine pleasure to use must not only be concentrated on avoiding negative aspects, but also strive to induce positive aspects (Jordan and Servaes, 1995). Properties that determine the pleasurability of a product include functionality, usability, aesthetics, interacton and emotion. The emotions engendered by using pleasurable products include security, assurance, confidence, excitement, satisfaction, and entertainment.

In agent design, new or increased usage of an agent on a website will be mainly related to its functionality or usability. No matter how good it looks, users are unlikely to continue using an agent if they find it difficult to use or it does not perform the function they require. However, if we assume that the usability and functionality are appropriate for the target users, the image or design of an agent may either encourage users to use the agent or discourage them. As mentioned above, an attractive and appealing "social artifact" may increase the enjoyment a user has when interacting with it (De Angeli et al., 2002). In addition, different designs may be more or less appealing to specific groups, but these may be made up of a different set of users than those envisaged by the design team. Thus, perhaps one of the solutions to this would be to consider the context in which the agent is being used and design it so that it is appropriate for the context.

Computer users want their interfaces to be more appealing and pleasurable to use (Nielsen, 1993), and therefore if an agent is to be used on an interface or website, then it too must be appealing to look at and pleasurable to interact with.

This thesis is concerned with investigating which type of agents people like, and the possible influence of aesthetics. It may be the case that agents need to be physically attractive to be pleasurable or appealing. On the other hand, people may prefer if the agent looks intelligent and smart, and physical attractiveness may not be the most salient characteristic in successful agent design.

It is therefore imperative that the appearance characteristics of an agent be part of the agent design process. Consequently, this thesis will investigate the appearance characteristics of agents that users prefer, and those that they consider most appropriate in agents. This will be done in a variety of different contexts, firstly in a financial setting, with further work investigating a number of different agent occupations and possible stereotyping. The contexts selected will be chosen to represent ecologically valid setting in which agents could potentially be used. In addition, the methodology used is drawn from standard psychological research. This will permit comparisons between this new applied empirical work and previous theoretical, purely research driven, findings.

Chapter 2 : Agents

2.1 Introduction

Many web sites have introduced computer agents on their home page to help users navigate around and to provide information about that particular company or service. Computer agents are also used on desktops, rather like a personal assistant, helping users surf the web, reminding them of appointments or reading their e-mail. A computer agent can be defined as "a bundle of functionality that performs some task for a person, either in real time or asynchronously....agents may be represented anthropomorphically but they need not be" (Laurel, 1993, p46).

Interface, or virtual, agents are not only becoming more and more popular on websites, but also in education and computer based training industries. One of the main principles of an interface agent is to motivate and encourage the user, and to provide assistance, either in learning tasks or on the Internet. Consequently, user perception and response to an agent may be crucial to their success.

When an interface agent is created, the process is an extremely complex venture that requires research in a diversity of domains, including A.I. (Artificial Intelligence), computer animation and graphics, interface design, sociology, and psychology (Xiao, Catrambone & Stasko, 2003).

While there have been several studies investigating the personification of agents (e.g. Koda and Maes, 1996; Cassell, 2000; Fielding, Fraser, Logan, & Benford, 2004), the research into interface agents has been dominated by the areas of animation (e.g. Clarebout, Elen, Johnson, & Shaw, 2002), anthropomorphism (e.g. Iwata, Takahiro, & Morihara, 1999; Isbister & Nass, 2000), and dialogue/interaction (e. g. Cassell & Vilhjalmsson, 1999; Cavazza, Charles & Mead, 2002). Although user centred design has been noted as being an important aspect of HCI (Laurel, 1990; Tannenbaum, 1998), there has been little consideration into user preferences or expectations in the design of interface agents.

2.2 User perceptions

A number of factors have already been identified as important to user perception. For instance, these include, the nature of the task, specific agent attributes, the characteristics of the user, and specific features of the task itself. These are discussed in more detail below.

2.2.1 The nature of the task

If the task required a great deal of interaction with the agent (such as a debate or conversation) the user may be more likely to attribute more human-like qualities to the agent. However, other tasks may lead the user to think of the agent simply as a reference tool (for example, aiding with a software system). One study that examined the role of the task being performed was carried out by Catrambone, Stasko, and Xiao (2002). They proposed a framework that not only emphasized the attributes of the agent, but also took into consideration the characteristics of the user, and the task they were performing.

2.2.2 Agent attributes

Catrambone et al. suggest that there are four main agent attributes to take into consideration: fidelity, presence, role, and initiative. Fidelity refers to the type of agent being used, whether it is animated (life-like), static, or iconic. Presence refers to the amount of time the agent is on the screen (i.e. is it always on the screen, or only when it is required?). Role refers to the role of the agent: is the agent is a partner, offering strategy guidance, or is it more 'low-level' and only offers 'how to' advice? Finally, initiative refers to whether the agent is proactive, offering advice or guidance, or reactive, only responding when asked a question. Other variables that could affect how an agent is perceived are how much expression it can show, quality of speech, gender and physical appearance of the agent, 'personality', and competence.

2.2.3 Characteristics of the user

In relation to the characteristics of the user, Catrambone et al. suggest that examining a user's personality traits may reveal how they will respond to agents, and how useful they will find them. In addition, background knowledge in a particular domain could influence whether a reactive or proactive agent would be preferable. A novice user may prefer a proactive agent who constantly offers advice, whereas an expert may prefer a reactive agent, and only receive help when they ask for it, usually at a low-level. Finally, other factors such as age, gender, and computer experience may affect how users perceive an agent.

2.2.4 Specific Features of the task

Finally, the features of the task a user performs with an agent can vary greatly, but Catrambone et al. suggest two main dimensions: 'Objectiveness' and 'Intent'. Objectiveness relates to the type of task being carried out. Tasks can be opinion-based, where the user is either seeking advice or recommendations or they can be objective, where the user is carrying out a task that requires finding out facts. Intent relates to the purpose of the task; the user could be a novice who is learning how to use a new system, thus requiring guidance, or they could be familiar with the domain and only want help with low-level details. Other variables that could affect perceptions of an agent are the domain in which the task is carried out, the extent to which time pressure is applied, the duration of the task, and the consequence of task performance.

Results of the study carried out by Catrambone et al. showed in an opinion-based task (travel advice), participants were influenced by the views of the agent regardless of whether it was life-like or iconic. In addition, the type of agent did not affect the time taken to complete a procedural task (text editing), perceptions of such attributes as how worthwhile, friendly, annoying, and intelligent the agent was, nor the enjoyment of interacting with the agent. Although there were no differences based on agent type, there was a general positive regard for the agents. The type of task had an effect on

two of the attributes: worthwhile and intrusive. In the procedural (editing) task, the agents were reactive and judged to be more worthwhile and less intrusive than the more proactive agents in the travel task. Catrambone and colleagues attribute this difference to the fact that agent was there to remind participants of the essential commands, thus making it seem more worthwhile while the agent in the other task criticised choices made by the participants and may therefore have seemed more intrusive. Interviews with participants revealed that, in both task conditions, the agents were considered as being helpful but in general participants interacted more, and more easily with, the editing agent than the other agent. Nevertheless, the agents in the opinion-based travel task were considered to have more 'personality' and to be more intelligent. Catrambone et al. suggest that although the editing task agents provided answers, the travel-task agents may have been seen as being having a deeper knowledge. This study not only highlights the importance of the type of task used when evaluating agents, but also the importance of the interaction between task type and whether the agent is proactive or reactive.

Xiao, Catrambone, and Stasko (2003) further investigated how the usefulness of an agent assistant influences performance on a novel text editing system compared to traditional help techniques (in this case it was a paper help manual), and how proactive behaviour from the agent affected performance and how the agent was perceived. The task in this study was to learn and use an unfamiliar text-editing tool. The three conditions in the experiment were paper manual, reactive agent (an agent that would respond to questions from the user), and proactive agent (an agent that would respond to questions as well as making additional suggestions). Due to anecdotal evidence about other proactive agents, such as Clippy, Xiao and colleagues predicted that the proactive agent in their study would have a negative impact on performance. The results of the study showed that objective performance of the tasks (e.g. time taken to complete the task, and the number of commands issued) was consistent across the three conditions. Subjective assessment results showed that there were no differences between the reactive and proactive agents on the attributes measured (these included how worthwhile participants considered the agent, how intelligent they thought the agent was, and how distracting it was). Participants in both agent conditions strongly felt that the help give by the agent was valuable, and the interactions with the agents were helpful. One of the most significant findings in

this study is that previous experience with, and stereotypes about, agents did not taint people's view of them, providing the assistance given by the agent was relevant and useful. However, Xiao et al. concluded that a user's preference for the different styles of help available, whether it be printed, on-line, or agent-based, will differ from person to person, and these preferences can be based on a variety of factors such as experience and the user's personality.

2.3 Attitudes towards the Use of Agents

In general, there are two main opinions in the use of agent-based characters in user interfaces; those supporting their use (e.g. Laurel, 1990; Cassell, 2000), and those questioning their use (e.g. Lanier, 1995; Shneiderman, 1997). However, there has still been little empirical research done in either area. Dehn and Van Mulken (2000) carried out a review of the empirical work done and their general findings suggested that there is still a great deal of inconsistency with regards to agents. Generally, they found that systems are perceived to be more entertaining when an agent is present compared to a system with no agent present. This does not apply to all systems however; systems generally considered to be 'attractive' are not affected by the addition of an agent. Similarly, there is variation with regards to the social perception of an agent. In some cases, life-like agents are perceived to be more intelligent that 2D images, in other cases there are no differences. The same is true for perceptions of attributes such as believability and usefulness. These inconsistencies can be partly explained if factors such as context and type of anthropomorphisation are taken into account, in addition to whether the user's perception of an agent is based solely on physical appearance.

Xiao, Catrambone and Stasko (2003) suggested that one of the main difficulties with carrying out empirical studies is that agents can in fact provide a whole host of different functions. For example, agents can be used as avatars, guides, intelligent assistants, or simply as entertainers on either the web or on the desktop. In addition to the functional aspects of an agent, they can also take on a profusion of different forms; they can be embodied or disembodied, personified, anthropomorphic, 2D, 3D, cartoon, static, interactive, video representations, or a combination of these factors.

The effect of agents is possibly exemplified by 'Clippy', the Microsoft Office Paper-Clip assistant, who is probably the most well known proactive agent, which many users have encountered. There is strong anecdotal evidence to suggest that many computer users have a negative view of Clippy, a view that is supported by a study by Xiao et al. (2003). They asked participants about their view on Clippy, and while all participants were familiar with Clippy, their impressions of it were explicitly negative, with most describing it as 'annoying'. However, while most views on Clippy were predominantly negative, some participants in the study by Xiao et al. (2003) claimed that they preferred the Office assistant when it was represented by the cat or dog. This suggests that the representation of an agent could have an influence on subjective reactions to it.

The effects of using interface agents have been studied in a number of task settings, such as interactive learning (e.g. Lester, Converse, Stone, Kahler, & Barlow, 1997), web browsing (e.g. Hook, Persson, & Sjolinder, 1999), and document composition (Rhodes, 2000). In the study by Lester and colleagues, an animated pedagogical agent, Herman the Bug, provided advice to students designing plants. Results showed that the proactive agent could increase not only the learning performance, but also the students' motivation. The active communication with 'Herman' had a strong positive effect on the students' perception of their learning experience, which resulted in them being more attentive and engaged in the task. Similar results were found in a study investigating the use of proactive help when using the COACH system, which helps users learning the Lisp programming language (Selker, 1994). When compared to the COACH system without proactive help, it was found that users in the proactive group, on average, made use of all available help materials, felt more comfortable with Lisp, had higher self-confidence, and wrote five times as many functions.

However, proactive help is not always a positive feature in an agent and in some cases has been considered as intrusive and offensive. Rickenberg and Reeves (2000) found users had a higher level of anxiety and made more mistakes when completing a task with a computer character that was monitoring their behaviour and responses, compared to a character that ignored them. However, further analysis of the results indicated that these findings were dependent upon the personality of the users, with

users who believed that others controlled their successes being more affected by the agent compared to those who believed they were responsible for their own successes.

There is a variety of factors that contribute to how positively an agent is viewed. The main finding in much of the research is that people are inclined to assign human-like characteristics to agents. Many researchers believe that anthropomorphic computer interfaces have immense potential to be beneficial for a number of reasons. As there is an ever-increasing amount of information in computer systems, and on the Internet, agents could be used as 'smart assistants', helping users to deal with this vast amount of material. Anthropomorphic interfaces could make a computer more:

- Human-like
- Engaging
- Entertaining
- Approachable
- Understandable

to the user.

This in turn may increase trust, and help to establish relationships with the user, thus increasing user comfort with computer (Catrambone et al., 2002).

Studies have in fact shown that anthropomorphic agents can be attention grabbing, and that people often attribute, and make assumptions about, the intelligence and abilities of these agents. The addition of simple factors such as eye blinking to a 3D character can result in increased levels of perceived intelligence, when compared to a non-blinking 3D character, a caricature, or geometric shapes (King and Ohyn, 1996).

However, the effect on attention can be either positive or negative it appears. Takeuchi and Nagao (1995) had participants engage in a conversation with an agent, in which a facial display was either present or absent. Results showed that the conversation where a face was present was more "successful", as it seemed to provide extra, important, conversational cues. However, these "successful" conversations required more effort from the user, and at times were found to be distracting. Conversely, Walker, Sproull, & Subramani (1994) found that the attention commanded can have a positive effect. Users filling in an online questionnaire spent

longer on it, made fewer mistakes, and wrote more commands when they were interacting with an anthropomorphic agent, represented by a talking face, compared to those who answered a text-based questionnaire.

2.4 Agent Disadvantages

These opinions are challenged by other researchers who consider anthropomorphic agent interfaces as being unrealistic and unsuitable. In addition to this, it is argued that current technology is still not advanced enough for natural language understanding, speech recognition, and learning capabilities to come even close to those of a human assistant. In particular, Lanier (1995) proposed that agent systems disempowered the user by making unclear such issues as who is responsible for the actions of a system. Others, such as Shneiderman (1997) believe that it is more beneficial (for users) that the interface commands are transparent and provide the user with objects that they can act upon. Additionally, critics argue that interface agents can be ambiguous for both users and designers, and may increase user anxiety, reduce user control, undermine user responsibility, and destroy a user's sense of achievement (Shneiderman and Maes, 1997). Anecdotal evidence also suggests that users find many of the current anthropomorphic agents, or personified interfaces, annoying, silly characters that hinder rather than enhance productivity.

2.5 Empirical Evidence

Although arguments have been made both for and against the use of interface agents, relatively little empirical research been conducted in this area. The research that has been done has produced quite contradictory or ambiguous results (Cassell, 2000; Dehn & Van Mulken, 2000).

Hook, Persson, and Sjolinder (2002) created the Agneta and Frida system in order to enhance the overall experience created by interacting with virtual characters. They envisaged that placing the two characters on the desktop would encourage the exploration of the information space. They also saw it as a way of mixing a narrative element into the mainly spatial experience of using the Internet. The general findings of the investigation into how users reacted to Agneta and Frida showed that

interaction induced positive as well as negative emotions in users. Although many of the users indicated that the characters made browsing a more relaxing experience, other users indicated that they were disturbed by Agneta and Frida, particularly those with a lot of web and/or computer experience. Hook et al. attribute this disturbance to factors such as experienced users having a strong model of web exploration, or a general dislike for interactive characters.

One of the main purposes of an interface agent is to motivate and encourage the user, and to provide assistance, either in learning tasks or on the Internet. Consequently, user perception and response to an agent is crucial.

2.6 Perceived Intelligence of Agents

There is mixed evidence on the ways in which agents affect a user's behaviour and attitudes towards an interface. The social perception (such as perceived intelligence and likeability) of anthropomorphized agents is, in some cases, is that they are more believable and useful than non-anthropomorphized agents; in other cases, the opposite is true. Koda and Maes (1996) found that participants perceived a caricature male to be more intelligent than a caricature dog, based solely on visualisations. However, if the subjects interacted with the system (in this case playing poker) visualisations did not affect the perceived ratings of intelligence, with an opponent who was not visualised and one who was visualised with a caricature face being rated as having the same level of intelligence. Sproull, Subramani, Kiesler, Walker and Waters (1996) found similar results when they asked subjects to rate the intelligence of a virtual counsellor on a career-counselling system. They found that there were no differences in the perceived intelligence of the counsellor when the subjects were presented with just written text or when an animated face spoke to them on the screen. However, they also found that in terms of social evaluation (e.g. attractiveness and friendliness), subjects rated the virtual counsellor higher if presented with written text that when presented with the animated face. Koda and Maes reported higher likeability in a poker game for an opponent visualised by a face rather than an invisible opponent. Dehn and van Mulken explain this inconsistency by the type of agent chosen: Koda and Maes used a 2D caricature male face whereas Sproull et al. used a realistic

animated 3D face which spoke (judgements may have been based on the voice used, not just the face).

Consistent with the findings of Clarke, Jordan and Cockton (1995), Takeuchi and Naito (1995) found that a virtual card matching game was perceived as more entertaining when an opponents' moves were visualised with an animated face as opposed to with an arrow. Koda and Maes (1996) also found that a poker game was judged to be more engaging when playing against a visualised opponent. It may be the case however that these findings are domain-specific. There is evidence to suggest that if the task or interface is visually attractive then the presence of an animated agent might not make any difference. Van Mulken, André and Müller (1998) (cited from Dehn and van Mulken, 2000) studied the difference between using an animated agent to introduce fictitious employees of a research institute against using an arrow. They found that the entertainment ratings of the two systems did not show any real differences. This could be attributed to the use of the photos of the employees, which may have made the interface entertaining or visually attractive from the start.

2.7 Effect of Agents

There are several factors that could influence the results in studying the effect of agents. The first of these is whether the physical appearance of the agents is the only information on which subjects can base their opinions. Secondly, the particular type of agent is chosen (whether it is anthropomorphised or not), and lastly the context or domain in which the interaction with the agent is set. These factors are particularly relevant when evaluating social attributes as some agents are thought to be more intelligent than others based on physical appearance. However, the effect of physical appearance may disappear if the system demonstrates characteristics that are more diagnostic of the feature to be assessed. For example, when playing poker against a computer system, subjects would tend to base the intelligence rating of the opponent on their poker playing skills rather than on their appearance.

Social evaluations, such as likeability and comfortability, are also dependent on what particular kind of anthropomorphization is shown. A fully animated 3D display of a

face will probably appear much more like a real person than a 2D caricature face, and hence may be evaluated on the basis of this criterion.

As agents are becoming more and more popular in education, computer based training, and on the internet, the impact of the physical appearance characteristics of the agent needs to be elucidated. While there have been several studies investigating the personification of agents, as already said, the research into interface agents has not taken into consideration user preferences or expectations with respect to the physical appearance of interface agents. This thesis aims to determine the impact of physical appearance, and establish if this interacts with the context in which an agent is used.

Chapter 3 : Social Psychology and Agents

Nass and colleagues have carried out numerous studies that examine how people react to, and interact with, computer systems and applications that include a variety of personified characteristics (e.g. Nass, Isbister, & Lee, 2000; Nass, Steuer, & Tauber, 1994; Rickenberg & Reeves, 2000). The primary finding in this work is that people treat and react to computers in a social manner, similar to how they personify other people and, therefore, they propose that findings in social psychology literature may apply even when one of the two participants is a machine. Some of the findings that may have particular salience to interactions with agents on websites are discussed below.

3.1 Person Perception

When we first meet other people we form impressions of them in several different ways. It has been recognised for some considerable time that implicit theories about personality traits are often used, whereby a person is 'assigned' a number of traits because they possess one particular central character trait. Describing a person using 'central traits' such as 'warm' or 'cold' can greatly influence our views of an individual, whereas other traits such as 'polite' or 'blunt' have been found not to be as important (Asch, 1946).

A number of theorists have suggested that certain personality traits are related to one another; that we have a generic way of characterising people and that variations in central traits influence how we view others. For example, Rosenberg, Nelson and Vivekanethou (1968) identified two major dimensions: Intellectual and Social, and proposed that knowing where one trait lies on these dimensions, one could assume that there were other characteristics associated with it. Kelly (1955) proposed an alternative theory, Personal Construct Theory. He believed that people develop their own theories about what others are like and that most people have different theories about what personality traits mean, as they are formed from their own experiences. However, the traits used to make a judgement need not always be personality traits.

For example, Maier (1955) found that people's job descriptions were enough for others to make judgements about what that person was actually like.

However, theorists generally agree on the importance of a number of factors, some of which are first impressions, attractiveness and stereotyping.

3.2 First Impressions

One of the most salient factors in perceptions of people is first impressions of others. The impressions we form about someone are not only dependent upon the information we receive – they are also based upon the order in which the information is given according to Asch (1946). Asch gave two groups of participants six adjectives describing a person. In the first group the positive adjectives were at the beginning of the list but in the second group the negative adjectives were first. Later, the first group rated the person more favourably than the second group, thus demonstrating the importance of first impressions, and the effect that a good first impression makes. However, it has been shown that the first time people meet, whether it is in a social situation or a work related situation, the characteristic that impresses the most is physical appearance (Cash and Janda, 1984). Impressions of women in particular are especially susceptible to opinions on their physical appearance according to Abramowitz and O'Grady (1991).

3.3 Attractiveness

An important contributory factor in the effects of physical appearance is attractiveness. There is a substantial amount of research that suggests that we are more positive towards people whom we find physically attractive than those we find physically unattractive (Hayes, 1993). Adams (1977) proposed that physical attractiveness effects are rarely determined by environmental influences, but that they are present across a diverse range of experience and ages, including attitudes of parents and others to children (e.g. Dion, 1973; Dion & Berscheid, 1974; Adams & LaVoie, 1975). In the case of adults, Goldman and Lewis (1977) had participants engage in telephone conversations with unseen strangers of the opposite sex. Results showed a positive correlation between judgements of social skills and attractiveness,

in spite of the fact that participants did not see the partner in their interaction. Goldman and Lewis (1977) suggest that this is due to attractive individuals having a different socialization experience when growing-up compared to unattractive individuals, thus having more confidence to portray themselves as being attractive.

The effects of attractiveness have also been shown to go beyond social skills judgements. For instance, in a study of college students and the effect of attractiveness on the evaluation of their work, Landy and Sigall (1974) concluded that "if you are ugly you are not discriminated against a great deal as long as your performance is impressive. However, should your performance be below par, attractiveness matters. You may be able to get away with inferior work if you are beautiful" (p 302).

3.3.1 Factors in Attraction

The general conclusion of much of the research on personality and physical attractiveness is that observers seem to apply a generally positive stereotype to persons who are physically attractive. Thorndike (1920) called this tendency the 'halo effect', whereby if a person has one positive trait then this tends to elevate ratings of other traits, even when there is little or no information to support this. This suggests that attractiveness is a universal construct and that it will influence people's judgements on other personality traits.

Berscheid (1985) suggested that our attraction towards others is really about our overall impression of the person, whether they are a generally 'good' or 'bad' person, which could explain the stereotype effect found by Dion et al. (1972). He suggested that attraction and dislike are the two main ways in which we decide on our responses to other people.

As to the factors that produce attraction, although it is thought that 'opposites attract', research has in fact shown that the more similar two people are, the more they will be attracted to each other. Byrne (1961) found that people liked others with similar attitudes significantly more than those who differed, as well as rating them as being

more intelligent. This effect is particularly prominent in first impressions, but seems to deteriorate with time. In relationships, for instance, engaged couples who had been together for more than eight months did not think that having similar attitudes to their partners was as important as those who had been engaged for a shorter time (Kerchoff and Davis, 1962).

Similar findings have been demonstrated in regard to facial attractiveness and similarity. Little, Burt, Penton-Voak and Perrett (2001) showed that the faces found to be most attractive are the ones that look most like us. Participants consistently found one face more attractive than the others that they were shown, and that face happened to be their own, but of the opposite sex. However, they warn that the results may not be as straight forward as indicating that people simply prefer others who reflect themselves. Since the images used in this work were altered in such a way as to produce more average features, but of the opposite sex, this 'averageness' rather than similarity may have been the factor causing them to appear more attractive. Studies by Langlois (Langlois & Roggman, 1990; Langlois, Roggman & Musselman, 1994) have in fact shown that people are generally attracted to average characteristics in a face. However, Perrett, May and Yoshikawa (1984) purport that while the average of a set of faces may be attractive, it is not the most attractive. In a study using digital composites, they found that the most attractive face was one that was the average of a set of attractive faces, when compared to the average of the whole sample of faces.

The link between similarity and attraction seems also to extend to product design. Research on Product Personality Assignment (PPA) has shown that products reflecting participants' personalities were significantly preferred over other products (Jordan, 1997b), which could have relevance to agent design if agents are regarded as products or product-like.

3.4 Stereotypes

The perception of attractiveness has repercussions, which go beyond the merely physical. Dion, Berscheid & Walster (1972) investigated two things, whether people employed physical attractiveness stereotypes, and, if so, the content of these

stereotypes. Their results showed that attractive individuals were judged to be more socially desirable than unattractive individuals, regardless of the gender of the rater. Attractive stimulus persons (whether male or female) were judged as having better occupations, being better partners/spouses and having happier marriages. They were also assumed to have better prospects for social and professional happiness. Dion et al. suggest that "physical attractiveness stereotypes do exist and that their content is perfectly compatible with the 'what is beautiful is good' thesis" (Dion et al., 1972, p289).

Most of the literature on attractiveness stereotypes has agreed that the beauty-is-good stereotype is a strong and general phenomenon. However, Dion (1981, 1986) suggested that the effects of the stereotype are strongest when making judgements about social competence and interpersonal ease. Bassili (1981) reached a similar conclusion that the strongest link with attractiveness was social vitality or extraversion. Using the components identified by Rosenberg (1977), Eagly, Ashmore, Makhijani and Longo (1991) carried out a meta-analysis to investigate whether physical attractiveness was strongly linked to any particular dimension of implicit personality theory. The sample of studies used only included those that addressed the beauty-is-good stereotype. Results showed attractive targets in the studies were evaluated more favourably than unattractive targets. However, again, the biggest difference between attractive and unattractive targets was on ratings of social competence, and the smallest on ratings of concern for others and integrity. Eagly et al. also predicted, on the basis of previous work, that there would be a greater stereotype effect for female targets compared to male targets. However, as in the Dion et al. (1972) study, this was not found to be the case. Eagly and colleagues suggest that one reason for this may be that strong stereotypical judgements might only occur in face-to-face interactions with females (or in situations where interaction is likely). Additionally, differences in male and female participants' perceptions of the physical attractiveness stereotype differed very little. However, it is worth remembering that such lack of gender difference was not found in the Abranowitz and O'Grady (1991) study into the effects of first impressions.

In an investigation of the relationship between attractiveness and judgements of intelligence, Ashmore, Tumia and Schreier (1980) (as cited in Eagly et al., 1991)

studied the variation in responses to a stimulus person when they used the adjective "good-looking" compared to when they used "unattractive". The responses indicated that the attractiveness manipulation had little influence on judgement of intellectual competence but again strongly influenced views on social competence.
Overall, there is a great amount of variation in the strength of the beauty-is-good stereotype in the studies considered and also from measure to measure within studies. The results are dependent on the type of judgements participants are asked to make – physical attraction seems strongly to affect judgements on social competence, but less strongly judgements about potency, adjustment, and intelligence, and has little impact on beliefs about integrity and concern for others.

However, the variation in the observed strength of the beauty-is-good effect (Dion et. al., 1972) could equally result from different opinions about what 'good' actually means within the context of each study. There is no obvious agreed definition of 'good', and may have meant different things in different contexts.

3.5 Gender Stereotypes

Research also suggest that one of the ways in which people judge others is by combining a variety of different stereotypical characteristics and then basing their judgement on the characteristics that are most salient in each category (Grant, Button, Hannah & Ross, 2002). In their study Grant et al. asked participants to rate pictures on five different issues (attitudes to discipline of children, homosexuality, feminism, immigration, and religion). The results of this study demonstrated participants were largely relying on stereotypical judgements when inferring other people's attitudes. When compared to men, women were considered more opposed to strict discipline, more open-minded and accepting of homosexuals and immigrants, more in support of feminism, and more religious. In addition, they found that older people were regarded as being more conservative on all five issues, and that attractive people were more liberal.

These results not only demonstrated that people tend to utilize the range of information available to them in forming their perceptions of others, but that inferences based on gender, for example, were qualified by age. For example, the

gender stereotype that women would be more supportive of feminism compared to men weakened as the age of the person (in the picture) increased, with participants perceiving older persons (male and female) to have relatively little support for feminism.

In general, women were considered to be less strict when disciplining children, but as the attractiveness of the woman in the picture increased, this trend weakened and eventually reversed. Conversely, attractive men were considered to have attitudes that were more liberal than less attractive men and, in some cases, more liberal than women. (Grant et al., 2002)

Grant, Button, Hannah and Ross (2002) suggest that the results of this study demonstrate that people use multiple trait information to make qualified judgements about others. In addition, they suggest that other attributes not measured (e.g. race, or status) may have equal influence, and that the context in which the person is being judged will determine which attributes are more salient.

In addition, other factors also appear to interact with gender and, indeed other stereotypes. For example:

3.5.1 Expectations

Cann (1993) has suggested that gender-consistent information is more accurately recalled than gender-inconsistent information but more importantly, that when someone is employed in a gender-consistent occupation people recall that person's competence. In contrast, when people are employed in a gender-inconsistent occupation people recall that person's incompetence. For example, sentences like "Jane is a good nurse" and "John is a bad nurse" were better remembered than "Jane is a bad nurse" and "John is a good nurse."

3.5.2 Environment

In a similar study, it has been suggested that people associate person type with the physical environment in which they are situated. Lawrence and Leather (1999) investigated whether different environmental conditions influenced the stability of the occupational stereotype of a licensee (pub manager). Although licensees in general

were perceived to be more confrontational than negotiative, the opposite was true if they were shown in the context of a tidy public house (i.e. the 'right' environment). Lawrence and Leather (1999) suggest that people feel more able to judge the behaviour and beliefs of others if they receive contextualized information about the person. Similarly, people will make more distinguishable judgements about a person and "dilute their stereotypical judgements" (p391) if they have sufficient personal information.

In a similar line of research, interpretations of facial emotions were found to be dependent upon the context of the situation (Clarke and Russell, 1996). Algoe, Buswell and DeLamater (2000) suggest that any contextual cue that elicits stereotypes can influence interpretations, and found significant effects of the influence of gender and job status on the interpretation of facial emotions.

3.6 Factors influencing agent suitability

3.6.1 Beauty is good

As indicated above, studies across a variety of disciplines and settings, ranging from courtrooms and classrooms to social interactions including dating and interpersonal attraction, and hiring and advertising, have shown that attractiveness influences people's judgements, attitudes and actions. Most of these findings have agreed with the "what is beautiful is good" paradigm (Dion et al., 1972). Similarly, market research studies have investigated the effect of using attractive models and endorsers in advertisements (e.g. Baker and Churchill, 1977; Caballero and Pride, 1984; and Joseph, 1982) and have provided further support for this hypothesis. The issue that now needs to be considered is whether and how interaction between with web-based agents is affected by the same factors. Given too, the influence of context on social judgements, the potential role of the agent is also likely to play a part.

3.6.2 Attractiveness and Sales Performance

One of the roles a virtual agent can adopt on a website is that of a sales person. However, there has been very little research carried out specifically regarding the

effect of attractiveness on sales performance. Churchill, Ford, Hartley and Walker's (1985) meta-analysis of the hundreds of studies on individual sales performance identified a range of personal factors thought to influence selling effectiveness (e.g. height, weight, gender, age, neatness, appearance, education, race, and marital status) but in general, only small effects were found. However, two studies that have specifically looked at the effects of physical attractiveness on sales performance (DeShields, Kara & Kaynak, 1996; Reingen and Kernan, 1993) found that physically attractive salespersons have a better sales record than unattractive salespersons. Ahearne, Gruen and Burke (1999) went on to investigate the factors that moderate the effect and in particular the role of the salesperson-customer relationship in that it has been found that the attractiveness effect weakens as the relationship increases. In addition, they addressed how the attractiveness effect worked; whether it was direct or whether it was moderated by other factors as previous theories suggest that it is not the salespersons' attractiveness that directly influences the perceptions of sales performance; rather it influences an aspect, such as the customers' impression of the salesperson's trustworthiness or likeability, and it is these factors that influence the ultimate perception of performance (Ahearne et al., 1999).

A review of the impact of attractiveness of persons promoting products found that, compared to unattractive spokespersons, attractive spokespersons were liked more, viewed more favourably, were more persuasive, and had a positive effect on products with which they were associated (Joseph, 1982). Similarly, Reingen and Kernan (1993) demonstrated that, in the personal sales arena, physically attractive salespeople were judged as having more appropriate selling skills, and consumers treated them more warmly. Purchase intentions have also been found to be influenced by physically attractive salespeople (DeShields et al., 1996).

3.6.3 Mediators of Attractiveness

The physical attractiveness effect has been found to be greatest on first impressions (Reingen and Kernan, 1993), but that it may not be long-lasting. In a review of studies on consumer behaviour, Rao and Monroe (1989) concluded that buyers make inferences about the quality of an unfamiliar product from concrete attributes such as

price, appearance, or the packaging. These attributes are used (because they are easy to judge and are familiar) until a buyer gains experience with the product. Using this rationale, Ahearne et al. (1999) propose that as the physical attractiveness of a salesperson is easy to judge, this judgement "may be used by the buyer to make inferences about the more abstract variables such as the quality of the product or the credibility of the source" (p272). As the salespersons-customer relationship lengthens, the customer is able to make judgements of the other attributes (Dwyer, Schurr & Oh 1987). In a comprehensive study into the mediators for the effect of attractiveness, Debevec, Madden and Kernan (1986) found that the two attributes that fully mediated the effect were the knowledge and trustworthiness of the salesperson. Similar studies have examined the mediators of likeability (Joseph, 1982), and trustworthiness and expertise (Ohanian, 1991).

However, neither salesperson nor product is physically present during web-based interactions and it is therefore difficult to predict the effects on perception of the product under these circumstances.

Ahearne and colleagues confirmed their prediction that the perceived attractiveness of a salesperson would have a positive effect on sales performance. They had further proposed that the communication ability, likeability, expertise (in their field), and trustworthiness of a salesperson would mediate the relationship between perceived physical attractiveness and performance (i.e. physical attractiveness would have a positive effect on these attributes, and they, in turn, would have a positive effect on sales performance). Analysis of the data confirmed the importance of all but one of these relationships - the relationship between expertise and sales performance. Ahearne et al. explain this finding as being attributable to the fact that the salespeople were rated similarly as having a very strong knowledge of their area on the basis of expectation that that would be the case. Therefore, other factors influenced discrimination to a greater degree. Further analysis revealed that the relationship between attractiveness and sales performance was not completely mediated by communication ability, likeability, expertise and trustworthiness. This suggests that there may have been other mediators influencing the effect of attractiveness that were not addressed in their study. 3.6.4 Attractiveness and Persuasiveness

One of the ways in which agents can be used on a company website is to be a 'voice' for the company. They can be seen as an extension to the 'brand' or image of the company and persuade users to buy from the company (if is it sales based) or to offer advice (if it is service based). By using an attractive agent, people may be more likely to take the advice of an agent, as research has shown that people are more likely to agree with the opinions of attractive adults than with unattractive adults (Horai, Naccari & Fatoullah, 1974). Research in marketing and advertising has shown that attractive communicators are frequently liked more and perceived more favourably compared to unattractive communicators (Joseph, 1982). Kahle and Homer (1985) found a positive relationship between physical attractiveness and persuasiveness in marketing communications. Consumer attitudes towards advertising and advertising brands can be positively influenced by a source or endorser who is likeable (or has celebrity status) (Belch and Belch, 1998). This supports earlier research that states that physical attractiveness significantly affects the persuasive power of advertised products (Baker and Churchill, 1977; Caballero and Pride, 1984).

In their study, Phau and Lum (2000) investigated the influence of physical attractiveness of endorsers on consumer purchase intentions of two products, a slimming programme (a high involvement product), and a skincare moisturiser (a low involvement product). Results showed that there were significant differences in purchase intentions for both the skincare moisturiser and the slimming programme depending on whether it was advertised using an attractive, unattractive, or no endorser. Purchase intentions were higher if an attractive endorser advertised the products, but there were still significant positive correlations found between unattractive and no endorsers and purchase intentions. Purchase intentions were more influenced for the high involvement product (slimming programme) than for the low involvement product (skincare moisturiser) when the endorser was attractive. However, Phau and Lum suggest that this stronger trend towards endorser influence could be due to the fact that the slimming programme was an unfamiliar product therefore consumers would want more information in order to make a choice on whether to purchase it or not. This corroborates the findings of Landy and Sigal (1974) who suggested that the influencing effects of endorsers are strongest for less established products with less discernible benefits. Another contributing factor to the results in this study is the fact that the two products used were beauty products, which

may explain why an attractive endorser was most influential. This would substantiate the suggestion made by Till and Busler (1998) who suggested that the endorser has to 'match-up' with the product being advertised. However, Phau and Lum state that an unattractive endorser may influence consumer purchase intentions, so long as the advertisement was message driven, and that the target consumer group was highly involved. Therefore, it may be an effective approach to use an "ordinary, unattractive, or unknown spokesperson" (p.54).

3.7 Limitations to the Attractiveness Effect

On the other hand, there has been some contradictory evidence regarding the efficacy of physical attractiveness across differing media and advertising situations. Caballero et al. (1989) found no effect of physical attractiveness in the advertising of grocery products in a naturalistic setting (i.e. in a grocery store with store customers). However, this may not be surprising, as the product was in no way related to attractiveness or beauty. Similarly, Baker and Churchill (1977) found interaction effects between the attractiveness of the model endorsing a product and the type of product (attractiveness-related or attractiveness-unrelated). They also stated that the impact of physical attractiveness might be restricted by the gender of the consumer, and the gender of the spokesperson (model/endorser). Thus, the attractiveness stereotype may be mediated across different contexts by interactions between the persuasion situation, product characteristics, consumer characteristics, and the gender mix of the endorser-consumer dyad. This suggests that the product, and whether it is attractiveness-related or gender-specific, together with the gender of the target audience/consumer, must be taken into consideration when investigating the effect of attractiveness. Ahearne et al. (1999) suggested that one of the possible reasons for the mixed results in these studies is due to their methodological limitations. Many of the previous studies in this area have been laboratory simulations, rather than real-world simulations, and have relied on students as proxy-subjects. They also failed to measure actual behaviours, but instead relied on behavioural intentions, the two of which have been found to be conceptually distinct (Bentler and Speckart, 1979, c.f. Ahearne et al., 1999)).

Consumers have been found to be more likely to accept a product's claims made in advertisements and change their attitudes and buying behaviour if they believed the advertisement to be trustworthy and credible (Goldberg and Hartwick, 1990). Belch and Belch (1998) suggested that, in general, physical attractiveness would have little impact on consumer preference when the source expertise is high. Conversely, when task-related source characteristics, such as expertise are weak, consumers rely on 'irrelevant' cues, such as physical attractiveness, on which to base their opinions, and consumers tend to agree with a highly attractive source compared to a medium or low attractive source (Joseph, 1982).

Many researchers suggest that in order for consumers to relate to an attractive source and believe that the advertisement claims are credible, an association between the source and the product must be established (McCracken, 1989). Likewise, Till and Busler (1998) suggest that this "match-up hypothesis" be extended so that there is a fit between the (celebrity) endorser and the product. This suggests that the 'best' choice of agent for a website might be one where it matched the perceived stereotype of the occupation it is portraying. One of the ways in which this 'match-up' can occur is by using gender appropriate agents.

3.8 Issues for agent design

If users do indeed treat computers as "social actors" (Reeves & Nass, 1996), then it is expected that many of the issues raised in the social psychology literature, in particular, the effects of first impressions, attractiveness and stereotypes, may play a major part in determining how an agent is perceived by users. If the "beauty is good" hypothesis is true, then the aesthetics (or attractiveness) of an agent will be critical to its success.

In an advertising context it may be the case that if a product has a strong brand image, or clearly observable benefits, then the physical attractiveness of the model promoting it would be unimportant (Joseph, 1982). If, on the other hand, the product was a non-established brand, then irrelevant cues such as the physical attractiveness of the model may help to influence purchase decisions, and help selling effectiveness. This could be one of the ways in which designers could decide on the physical appearance of an

agent for use on a website. If the company using the agent is well known, the attractiveness of the agent may be of less significance than if the company is new and less well known. An attractive agent may help to promote the image of a company and could influence users to return to the website.

This, however, contradicts findings in the HCI literature that suggests that if a system is lacking in ergonomic qualities (i.e. usability), this can be compensated by an increase in hedonic qualities such as aesthetics (Hassenzahl et al., 2000). Similarly, Jordan (1999) suggests that the issue of pleasure (including aesthetics) in design is becoming increasingly important to users as they assume that products will be easy to use. Thus, if an agent was being used only as an extension to the advertising for a particular company, then how well know the company was could determine how attractive the agent would have to be. However, as agents are mainly used on a website for users to interact with then, so long as they are easy to use, aesthetics may be the most important feature in determining how well they are accepted or liked.

There are many reasons why advertisers frequently use pictures in advertising. These include attracting attention, enhancing consumers' memories for product-relevant information, aiding information retrieval, enhancing product judgements, influencing brand image, and attaching personalities to brands (Mitchell and Olson, 1981). Pictorial information has also been found to produce more favourable consumer attitudes compared to verbal information (Mitchell and Olson, 1981). Thus, effective communication can be established with consumers through a conscientious use of visual components in advertising. The same may be true for using virtual agents on a website. Agents may attract attention to a website and, in turn, users may spend longer browsing the site and interacting with the agent. Although agents are becoming more popular on the Internet, there are still only a number of sites that actually use them. By having a particularly well-designed agent on their website, a company can attract attention to their site and may also increase the likelihood of a user remembering the company. The agent may also become a personality for the company and induce positive regard towards the company.

The literature reviewed so far suggests that there are two main themes relating to this thesis: attractiveness and the 'beauty is good' hypothesis (e.g. Dion et. al., 1972), and the pleasure of interaction (e.g. Jordan, 1999). The following empirical chapters will

investigate these themes with a progressive set of experiments with the aim of teasing out the underlying relationships.

Chapter 4 : Aesthetic influence of agents

4.0 Aims of Chapter

Studies have suggested that one of the most salient factors in making judgements about others is attractiveness (e.g. Dion, 1973; Dion & Berscheid, 1974; Landy & Sigall, 1974). This may have implications for using computer agents: if an agent is 'aesthetically appealing' then any inadequacies or poor performance may be overlooked. However, it may also have the opposite effect – if a computer agent is 'good looking' then expectations about it's performance may be higher that if it were 'average' looking. This chapter examines aesthetic and utilitarian judgements of computer agents, based upon their visual image alone, presented as suitable for use within the context of a financial systems company. More specifically, the question being addressed is whether the social psychological constructs applied to people are equally applicable in interactions with computer agents.

4.1 Introduction and current study

A great deal of the research on stereotypes and the "beauty is good" hypothesis could have implications for designing virtual agents. If an agent is used on the Internet, how the agent looks could influence other judgements made. Thus, if an agent is attractive it may be thought of in more sociably desirable terms compared to if it were unattractive, and this could influence how users react to it and determine whether they would want to interact with it at all on a website.

Based on the work of Reeves and Nass (1996), it is predicted that many of the findings in this Chapter will mirror those in social psychological research on impression formation, with participants drawing upon stereotypical or schematic information sources to categorise the stimuli and make inferences about the agent's attributes, which cannot be inferred a priori from the image. In addition, Product Personality Assignment (PPA) research has indicated that products that reflect the user's personality are preferred over other products. Thus, an agent that is perceived as having a similar personality to the target user may be preferred over other agents, and hence may be the best choice for use on a website. Therefore, this study will also

investigate whether users ascribe similar personality traits to agents that they like to match their own personality, and, conversely, whether they ascribe a different set of personality traits to their own to agents they do not like.

The five main aims of this study are to determine:

- How people categorised a given set of agents;
- If people have a preference for a particular type or design of agent (male, female, non-human, or cartoon-like);
- Whether these preferences were influenced by aesthetics or by some underlying assumptions based on appearance;
- If personality attribution based on the physical appearance of an agent occurred and if so, whether people attribute the same personality traits to themselves as they do to their preferred agent;
- Whether people stereotype or apply the same rules to agents as they do to real people.

The selection of agents excluded well known characters where there might exist a pre-formed judgement about the character. Thus, characters such as 'Einstein', 'Shakespeare', 'Santa', and 'Merlin' were omitted. In addition, names were not given as research has indicated mixed effects (both positive and negative) as a function of name. It was felt that this would be a confounding factor as a person's first name may be one of the first pieces of information we receive about someone, and this can influence judgements on attractiveness (Erwin, 1993), as well as academic performance (Erwin and Calev, 1984; Harar and McDavid, 1973). This might suggest that there is a halo effect, whereby if a person had a desirable first name then participants would assign other desirable or positive attributes to that person/agent. On the other hand, unattractive names have also been found to be associated with a variety of positive traits. For example, Erwin (1999) found that individuals with unattractive names performed better on examinations and coursework at University compared to those with attractive names. He suggests the reason for this may be that the stereotype associated with names may be internalised and those with unattractive names may be spurred on to perform better than they would have otherwise.

The fourteen agents chosen included human (picture-like) agents, human-like (drawn) agents, cartoon human (caricature) agents, cartoon characters, and cartoon objects. These were selected from a population of around 150 agents currently in use or available for use on the Internet, and were selected as they encompassed the wide range of agents that are represented on the Internet.

The methodology used in this research is adapted from work carried out by Schenkman and Jonsson (2000), assessing the aesthetic appeal of web pages. The paradigm uses two stages of analysis, the first uses data to infer clustering of agent type and the second stage examines the predictive power of the clustering in relation to attributes along other critical dimensions. The intention is to collect pairwise judgements of similarity and preference scores and use Multidimensional Scaling to analyse the similarity data, in order to create clusters of agent type. The overall aim is to determine if the clustering of agents determined by MDS analysis, or from the preference data, predicts scores along dimensions such as attractiveness. For example, gender may be a dimension along which the agents were segregated and gender assignment may be used to infer other characteristics of the agent.

In this study, participants will be asked to make similarity and preference pairwise judgments, and then to rate each agent individually on a set of ten attributes (e.g. beauty, intelligence, and appropriateness). In addition, participants will also be asked to complete and Eysenck Personality Inventory (EPI) for the agent they liked and disliked the most, and for themselves. A within subjects design is employed in this study. The independent variables are the fourteen different agents used. The dependent variables are the scores on the similarity and preference scale, the ratings of each agent on each of the attributes measured, and the scores on the EPI. It is hypothesised that ratings of attractiveness would exert effects on judgements of other attributes. In addition, it is hypothesised that participants would ascribe personality traits similar to their own to agents that they preferred the most.[4]

[4] Some of this data was previously presented at CHI 2002 (Wilson, 2002)

4.2 Method

4.2.1 Participants

16 participants took part in this study, 8 were male and 8 were female. They were told that they were rating interactive interface agents to be used by a financial company on their web site. All participants were either staff or students of the Abertay University, and were selected by means of convenience sampling.

4.2.2 Materials and Apparatus

14 different agents were used as stimuli (see Appendix 1). The experiment was run on a standard desktop PC using SuperLab Pro for Windows, version 1.04. Instructions for task 1 described the purpose of the study and instructions on how to make the similarity and preference judgements for each pair of agents shown, along with an example of the scales that were to be used (see Appendix 2). Instructions for task 2 informed participants that they would be shown one agent at a time and the task was to rate each agent on a 10-point scale on ten attributes (these were beauty, pleasantness, appeal, intelligence, trustworthiness, sensibleness, usefulness, appropriateness, memorability, and overall impression) (see Appendix 3). The Eysenck Personality Inventory (EPI) self-report questionnaire was used to obtain a set of personality scores for each participant, and for the agents participants liked and disliked the most (see Appendix 4).

4.2.3 Procedure

The study involved 3 main tasks (described below), with the overall time taken to complete the study ranging from 45 minutes to 80 minutes, approximately. Task 1 involved making 182 pairwise comparisons and judgements and took approximately 20 minutes. In task 2 participants had to rate each of the 14 agents on 10 dimensions, resulting in 140 judgements, and took approximately 20 minutes. Participants had to complete 3 Eysenck Personality Inventory (EPI) questionnaires in Task 3 (57 questions on each questionnaire), which took approximately 20 minutes.

Participants first completed a set of practice trials for Task 1 (with a set of agents not used in the main experiment).

Task 1

Participants were given instructions for Task 1 to make pairwise comparisons and judgements between the agents. 14 agents were shown in pairs, with each agent being paired with every other agent. Each pair was shown twice, first to be rated on similarity and then on preference. This resulted in 91 pairs for similarity and the same number for preference (182 pairs in total) (see Appendix 1 for an example).

For the similarity judgement the endpoints on the 10-point scale were 'Fully dissimilar' and 'Fully similar'. For the preference judgements the 10-point scale ranged from 'Left hand agent preferred completely' to 'Right hand agent preferred completely'. The participants made their judgements by responding with the appropriate number on the keyboard. There was no time limit to make the judgements, and the next pair only appeared after participants had responded to the pair on screen. The agents were presented randomly so that possible order effects were balanced out. The agents were approximately 3"x2" in size, and were displayed on a 1024x768 resolution screen. They were shown in the middle of the screen, side by side, with word "Similarity" or "Preference" displayed above the agents, and the ratings scale below the agents. A grey square was displayed as a distracter stimulus for 500ms between each pair of agents.

Task 2

The second task was to judge each of the agents on a set of ten dimensions, including beauty, intelligence and overall impression. A set of instruction/answer sheets was given prior to beginning the task. Participants were asked to rate each agent on each of the 10 dimensions, along a 10-point bi-polar scale. The agents were presented, in a random order, individually on a PC. There was no time limit to make the judgements. The judgements were made on a separate questionnaire for each agent, and participants were instructed to write down the number of the agent they were rating.

Task 3

The final part of the main experiment required participants to complete an Eysenck Personality Inventory (EPI), which consisted of 57 yes/no questions. They were then

shown all 14 agents together and asked to choose which one they liked most and which one they liked least. Participants were then asked to fill in an EPI for the agent they liked most and the agent they liked least.

4.3 Results

4.3.1 Task 1

The similarity judgements were analysed using Multidimensional Scaling (MDS) (which allows analysis of any kind of distance or similarity matrix). Analysis showed that using two dimensions, 79% of the variance was explained (calculated by $\frac{\lambda_1^2 + \lambda_2^2}{\Sigma\lambda^2} * 100$, where λ_n is the n^{th} eigenvalue). The MDS solutions were interpreted by visual inspection (see Figure 1). The first dimension was interpreted as "human-ness", with the cartoon agents at one end and the human looking agents at the other end. The second dimension was gender, with a separation between male and female agents, and the cartoon agents in-between.

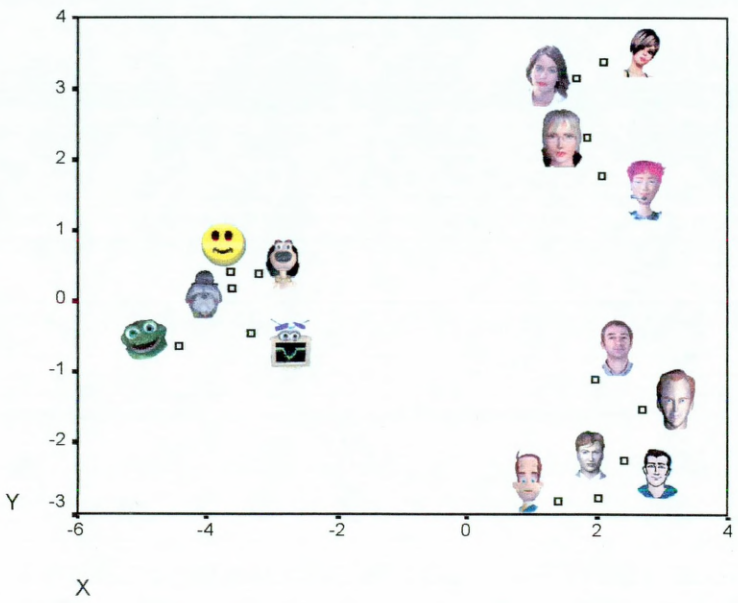


Figure 4.1: Similarity judgements - Plot in 2 Dimensions

Analysis of the preference data showed that the most preferred agent was Agent M (see over), who was preferred 100% of the time over each of the other agents. Overall, the female agents were preferred over the male agents and the cartoon agents.

Table 4.1

Percentage of judgements in favour of female agents in terms of expressed preference

Agent	% in favour
M 	100
C 	84.6
B 	76.9
E 	61.5
Average	**80.77**

Table 4.1, above, shows that participants preferred agent M 100% of the time when compared to other agents. In addition, it can be seen that female agents, in general, were preferred over 80% of the time when compared to other agents. The least preferred female agent, agent C, was still preferred 61.5% of the time over other agents.

Table 4.2

Percentage of judgements in favour of cartoon agents in terms of expressed preference

Agent	% in favour
I 	84.6
H 	76.9
J 	38.4
K 	23.1
G 	7.7
Average	**46.15**

Table 4.2, above, shows that agents I and H were the most preferred cartoon agents, with participants preferring them 84.6% and 76.0% of time over other agents,

respectively. In addition, it can be seen that cartoon agents, in general, were preferred just less than half of the time when compared to other agents. The least preferred cartoon agent, agent G, was only preferred 7.7% of the time over other agents.

Table 4.3

Percentage of judgements in favour of male agents in terms of expressed preference

Agent	% in favour
D 	53.8
L 	30.8
F 	30.8
A 	23.1
N 	7.7
Average	**46.15**

Table 4.3, above, shows that agent D was the most preferred male agent, with participants preferring him 53.8% of time over other agents. In addition, it can be seen that male agents, in general, were preferred less than 30% of the time when compared to other agents. The least preferred cartoon agent, agent N, was only preferred 7.7% of the time over other agents.

Table 4.4

Percentage of judgements in favour of each category of agents in terms of expressed preference

Preferred Agent – Comparison Agent	% in favour of preferred agents
Female - Male	100
Male - Female	0
Female – Cartoon	80
Cartoon – Female	20
Cartoon – Male	64
Male - Cartoon	36

Table 4.4, above, shows that when compared to male agents, the female agents in the selection were preferred 100% of the time, and cartoon agents are preferred 64% of the time. When compared to cartoon agents, female agents are preferred 80% of the time.

4.3.2 Task 2

In order to test for any significant differences between male and female participant ratings on the category scales (which were beauty, pleasantness, appeal, intelligence, trustworthiness, sensibleness, usefulness, appropriateness, memorability, and overall impression), a Mann-Whitney test was carried out and it was found that there were no significant differences between male and female participants on any of the 10 category ratings (see Appendix 5). Thus, it was decided to combine the data for male and female participants and report overall ratings.

A summary of the highest and lowest mean scores for each attribute is shown in Table 5 (see Appendix 6 for all mean overall scores). It can be seen from Table 4.5 that there was a general positive regard for female agents. Agent M was rated as more beautiful, pleasant, appealing, memorable than any of the other agents, as well as being rated highest in terms of overall impression. Agent C was rated as the most

intelligent, trustworthy, sensible, useful and appropriate agent. Agent K was judged to be the least beautiful, trustworthy, useful, appropriate, and memorable of all the agents, as well as being rated lowest in terms of overall impression.

Table 4.5

Highest and Lowest Mean Scores on Each Attribute

Agent	 C	 G	 K	 M	 N
Beauty	4.06 (1.73)	7.60 (2.66)	**8.10** (1.60)	**2.06** (1.06)	7.69 (2.06)
Pleasant	3.38 (1.50)	6.56 (3.03)	7.31 (2.27)	**2.38** (1.63)	**7.56** (2.45)
Appeal	4.00 (1.86)	6.94 (3.04)	7.75 (2.35)	**2.19** (1.64)	**8.12** (2.09)
Intelligence	**3.06** (1.12)	**7.88** (2.36)	8.50 (1.32)	4.50 (2.25)	5.31 (2.52)
Trust	**3.12** (1.78)	**5.94** (2.54)	**7.44** (2.58)	4.50 (1.23)	5.81 (2.59)
Sensible	**2.62** (1.25)	**8.06** (1.39)	7.75 (2.14)	4.56 (1.71)	5.00 (2.73)
Useful	**2.81** (1.17)	7.50 (1.93)	**8.50** (1.26)	3.56 (1.79)	6.06 (2.64)
Appropriate	**2.62** (1.26)	8.19 (2.20)	**8.56** (1.67)	4.75 (2.35)	5.88 (2.58)
Memorable	5.06 (1.61)	4.81 (2.29)	**6.50** (2.78)	**2.94** (1.91)	5.75 (2.93)
Overall Impression	3.94 (0.93)	7.31 (2.72)	**8.38** (1.31)	**3.81** (1.87)	7.44 (2.31)

- Highest score in category

- Lowest scores in category

A multiple regression was performed to determine the best predictors of the overall impression of an agent. The model predicting Overall impression included nine predictors (beauty, pleasantness, appeal, intelligence, trustworthiness, sensibleness, usefulness, appropriateness and memorability), and accounted for a significant amount of variance [adjusted R-square = 0.728; $F(9, 223) = 67.34$, $p<0.0001$]. The predictors Beauty ($t=3.52$, $p<0.001$), Appeal ($t=5.18$, $p<0.0001$), Intelligence ($t=2.14$, $p=0.033$), Appropriateness ($t=2.20$, $p=0.029$), and Memorability ($t=2.12$, $p=0.035$) all had a significant impact on predicting the overall impression of an agent. The impact of the predictors pleasantness, trustworthiness, sensibleness, and usefulness were not significant predictors in this model. The most important predictor was Appeal ($\beta=0.343$), followed by Beauty ($\beta=0.221$).

A principal component analysis was performed in order to further understand the relations of the category scales to each other. The analysis showed that beauty, pleasantness, appeal and overall impression were grouped together, indicating the importance of aesthetics/attractiveness for the overall impression. It also suggested that sensibleness, intelligence, appropriateness, usefulness and trustworthiness were grouped together, with memorability not falling into either of these two clusters (see Figure 4.2).

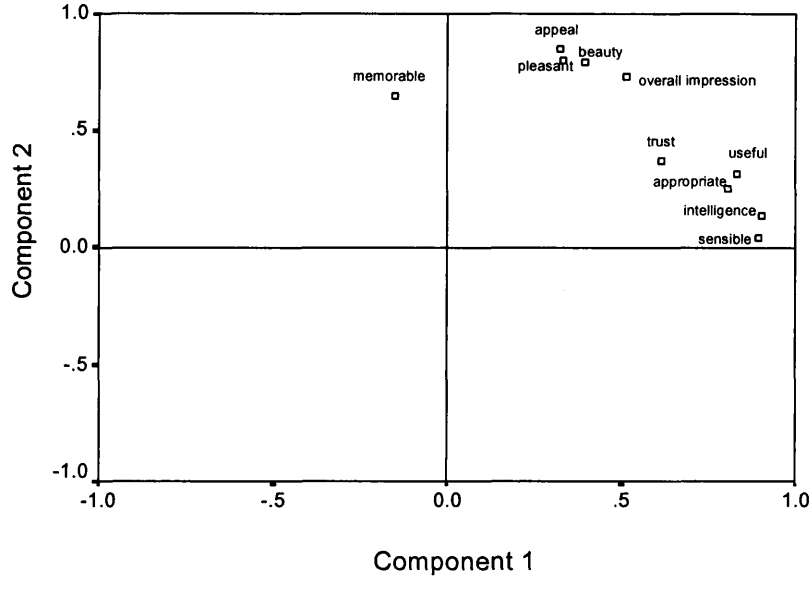


Figure 4.2: Factor loadings for category scales of Factor 2 plotted against those of Factor 1

4.3.3 Task 3 (EPI)

In order to test for any participant gender differences, six separate independent samples t-tests were carried out and it was found that there were no significant differences between male and female participants on any of the EPI scores (see Appendix 7 for statistical summary table). Thus, male and female data were analysed collectively.

Figure 4.3 shows participant's mean extraversion and neuroticism scores, as well as the mean extraversion and neuroticism scores for the agents that were liked most and the agents that were liked least.

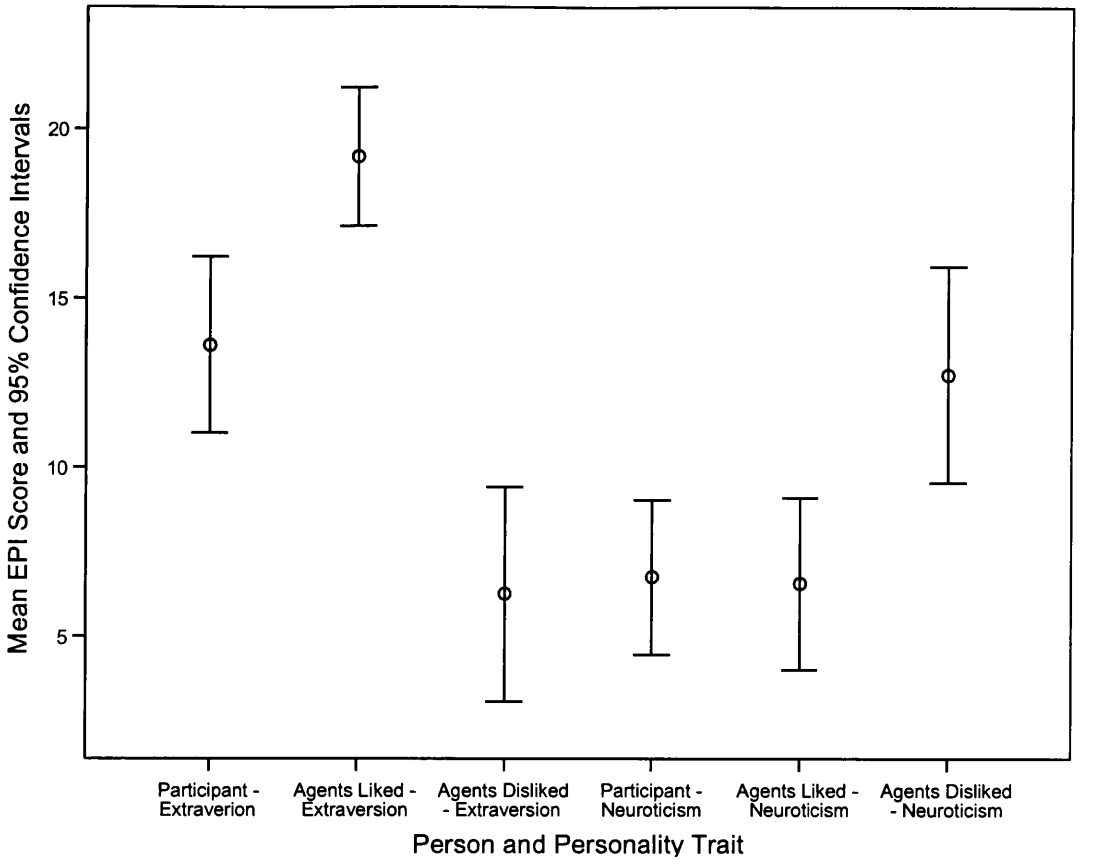


Figure 4.3: Extraversion and Neuroticism score for participant, agents liked, and agents disliked

Figure 4.3, above, shows that the mean Extraversion score for participants was 13.62, with the 95% confidence level falling between 11.02 and 16.23, the mean Extraversion score for agents liked was 19.19, with the 95% confidence level falling between 17.14 and 21.24, and the mean Extraversion score for agents disliked was 6.25, with the 95% confidence level falling between 3.08 and 9.41. The mean Neuroticism score for participants was 6.75, with the 95% confidence level falling between 4.47 and 9.03, the mean Neuroticism score for agent liked was 6.56, with the 95% confidence level falling between 4.01 and 9.11, and the mean Neuroticism score for agents disliked was 12.75, with the 95% confidence level falling between 9.55 and 15.95.

In order to test the observed differences between the participant's extraversion and neuroticism scores and the scores for the agents that they liked most and liked least, a

3 x 2 within subjects ANOVA (Person[5] x Personality Trait) was carried out and it was found that there was a significant main effect of Person [$F(2,30) = 4.241$, $p < 0.05$], a significant main effect of Personality Trait [$F(1, 15) = 15.168$, $p<0.05$], and a significant interaction between Person and Personality Trait [$F(2,30) = 34.231$, $p<0.001$].

Bonferroni pairwise comparisons showed that there was a significant difference between participants scores and the agents liked scores ($p < 0.01$) (Bonferroni corrected $\alpha=0.016$). No other pairwise comparisons were significant.

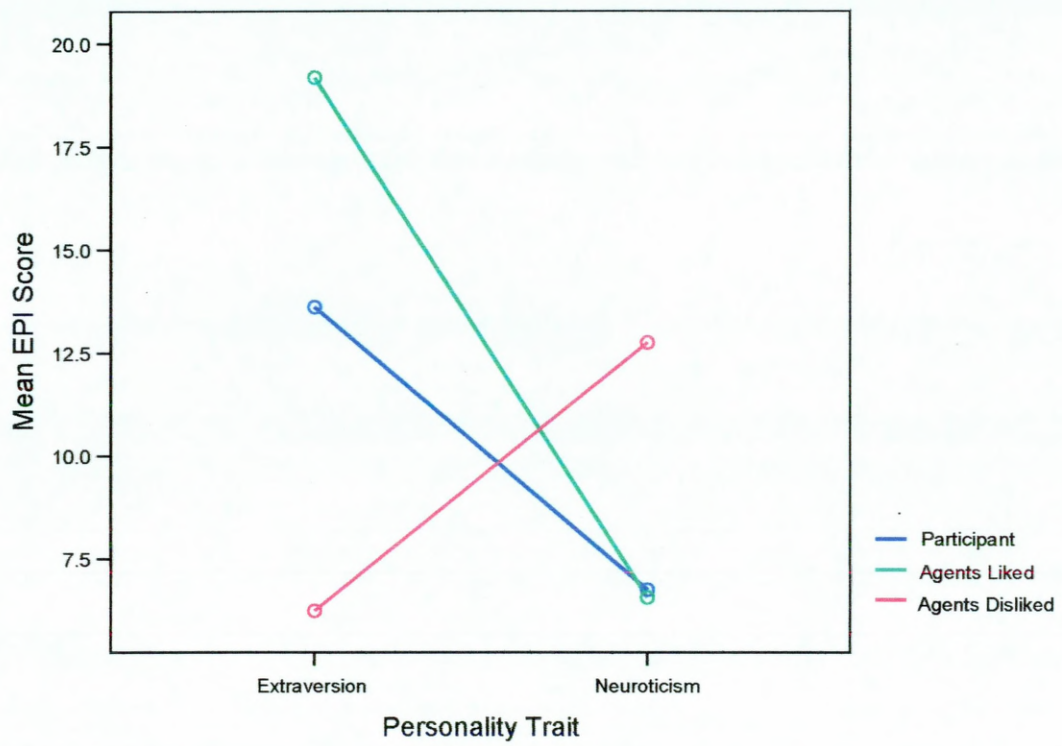


Figure 4.4: Interaction between Person and Personality Trait on EPI Scores

Figure 4.4, above, shows that participants rated themselves as being less extraverted than the agents they liked (13.62 v 19.19), but more extraverted than the agents they disliked (13.62 v 6.25). In addition, participants rated themselves and the agents they liked similar on neuroticism (6.75 v 6.56), but the agents they disliked were rated as being much more neurotic (6.75 v 12.75).

[5] "Person" refers to who participant's were scoring the EPI for (i.e. themselves (participant), the agent that they liked most, and the agent they disliked most).

In order to follow up the significant interaction between Person and Personality Trait, two 1 x 3 within subjects ANOVAs were carried out to test the difference between 'Person', firstly on Extraversion scores, then on Neuroticism score. It was found that there was a significant effect of Person on Extraversion scores [$F(2, 30) = 24.158$, $p<0.001$]. Bonferroni pairwise comparisons showed that there were significant differences between participants scores and the agents liked scores ($p = 0.001$), and the agents liked and agents disliked scores ($p < 0.0001$). The difference between participant's scores and the agents disliked score approached significant ($p = 0.021$) (Bonferroni corrected $\alpha=0.0166$).

It was also found that there was a significant effect of Person on Neuroticism scores [$F(2, 30) = 10.650$, $p<0.001$]. Bonferroni pairwise comparisons showed that there were significant differences between participant's scores and the agents disliked score ($p = 0.007$), and the agents liked and agents disliked scores ($p = 0.004$) (Bonferroni corrected $\alpha=0.016$). The difference between participant's score and agents liked score was not significant.

In addition, 3 separate paired samples t-test were carried out to test for difference between extraversion and neuroticism scores for each 'person'. It was found that there was significant difference between extraversion and neuroticism scores for participants [$t(15) = 3.571$, $p = 0.003$], agents liked [$t(15) = 10.283$, $p < 0.0001$], and agents disliked [$t(15) = -3.200$, $p = 0.006$] (Bonferroni corrected $\alpha=0.016$).

4.4 Discussion

The multidimensional scaling analysis of the similarity data revealed that there were two main dimensions. Participants discriminated between agents according to gender and "human-ness". Analysis of the preference data indicated that female agents were preferred over both cartoon agents and male agents. 100% of the comparisons involving male and female agents were in favour of females, and 80% of the comparisons between cartoon and female agents were in favour of females. By comparison male agents were only preferred 36% of the time over cartoon agents. Tables 4.1 – 4.3 show that the most preferred agent was Agent M who was preferred

over all of the other agents and selected, 100% of the time in preference comparisons. The next most preferred agents were Agent I and Agent C, who were both preferred 84.62% of the time. However, if looking at the scores for Agent A and Agent N (both male agents), their percentage scores in preferential comparisons were very low, 23% and 8%, respectively. A number of participants made reference to a 'maniacal expression' on Agent N's face and this may suggest that salient features or characteristics of the agent's image drive expressed preference. In particular, it may be characteristics other than attractiveness or beauty that disqualify an agents' persona with regard to preference data. It may be that categorical stereotypes, salient features, or population stereotypes drive judgements about an agent.

One would expect on average that 50% of the paired comparisons for preference would be for or against an agent by chance. Any agent with a score in the range 35% to 65% would not differ significantly from chance (i.e. choice would not be influenced significantly by their observable image characteristics). The preference data suggests that female agents would represent a more positive choice over cartoon or male agents. However, examination of the male agent scores suggests that none of the male agents examined were highly preferred and one conclusion could be that the sample of male agents used in the study was too restricted, and therefore the sample may have been biased with a high proportion of unattractive males. Alternatively, the sample may have been adequate and the effect found was real – that is, males were much less preferred compared to female and cartoon agents.

From Table 4.5 it can be shown that Agent M was rated as being the most beautiful and most appealing out of all the agents, and she also had the highest rating of Overall Impression. This supports Berscheid's (1985) theory that attraction towards others is related to our overall impression of them, and that there was a holistic overall impression that influenced judgements on other attributes. However, the attractiveness of an agent played a large part in determining peoples' perceptions of them. It may be the case that initial impressions of attractiveness created a 'halo effect' and lead to positive influence on other critical judgements.

Multiple regression analysis revealed that the best predictor for overall impression was appeal, followed by beauty. The regression analysis indicated that intelligence,

appropriateness and memorability had a significant impact on the overall impression of an agent. While these results seem to support research reported in social psychology literature, which suggests that we act more favourably to people whom we find attractive (Hayes, 1993), it is impossible to establish which dimension influences the others. Thus, it may be judgements about perceived intelligence, and not attractiveness, which influence other dimensions. These findings suggest that our overall impression of a person might be influenced by how attractive that person is, but there may be many other salient features that significantly influence the participant's overall impression of the agents.

Principle Component Analysis revealed that the rating of the category scales were grouped in such a way that appeal, beauty, pleasantness and overall impression were grouped together, and sensibleness, intelligence, usefulness and trust were grouped together, with memorability not falling into either clustering. This reinforces the idea that our overall impression of agents is initially dependent on their physical appearance. It also suggests that stereotyping also occurs when making judgements about computer agents – agents who were thought to be intelligent were also thought to be sensible, useful and trustworthy.

Task 3 was used to identify whether participants ascribe certain personality traits to agents that they either like or dislike, and whether these judgements reflect their personality. Although initial analysis revealed that there was a significant effect of 'Person' (irrespective of personality trait), it was only the difference between participant's scores and the agents liked scores that was significant when the pairwise comparisons were investigated. In this case, participants rated the agents liked significantly higher than themselves. There was also a significant difference between extraversion and neuroticism scores (irrespective of 'person'), with extraversion scores being significantly higher than neuroticism scores. Finally, there was a significant interaction between 'person' and personality trait. Further analysis of the significant interaction revealed that there was a significant effect of 'person' on extraversion scores, with the agents liked being rated as significantly more extraverted than participants. However, participants rated themselves as being significantly more extraverted than the agents disliked. In addition, there was a significant effect of 'person' on neuroticism scores, with the agents disliked being rated as significantly

more neurotic compared to participants and the agents liked. There was no significant difference between participant's scores and the agents liked scores. Furthermore, each 'person' had a significant difference between extraversion and neuroticism scores. Participants rated themselves and the agents liked as being significantly more extraverted than neurotic. However, the agents disliked were considered to be significantly more neurotic than extraverted.

Therefore, these results support the hypothesis that participants would ascribe personality traits similar to their own to the agents that were liked the most. This supports the research into Product Personality Assignment (PPA) (Jordan, 1997b), where products that are liked were judged to be of a similar personality to the user. Furthermore, participants judged the agents that they disliked to have significantly different personality traits compared to their own.

Many of these results support the notion that 'what is beautiful is good' (Dion et al., 1972). The overall impression of an agent was most influenced by aesthetic measures of beauty, appeal and pleasantness, all of which had a positive correlation with other attributes such as intelligence and how useful participants thought the agent would be. The attributes of the agents were also grouped together, in a manner similar to how people judge other people, with intellectual attributes clustered together and aesthetic factors clustered together.

Clarke, Jordan and Cockton (1995) demonstrated that users found a system to be more engaging if an agent was used, rather than just hypertext. The results in the present study, however, show that it is important to consider the type of agents users like and also what other attributes they assign to an agent based on it's physical appearance. Beauty, or aesthetics, plays a large part in determining how people perceive agents. If an agent is thought to be attractive, then this appears to influence our perceptions on other attributes such as how useful it might be, and how intelligent it is. There is, however, a danger in making an agent 'too attractive'. In the current study, Agent M was rated as being the most attractive, but average for intelligence, whereas Agent L was rated as being quite unattractive, but more intelligent. Overall, there was a positive correlation between beauty and intelligence however, which suggests that perceived intelligence is still related to how attractive an agent is. These findings

suggest that while beauty can be an influencing factor in how intelligent an agent is perceived to be, there may be more salient features, for example age, which also drives our judgements of intelligence.

Although the current experiment did not require participants to interact with the agents, much of the research concerning agents has involved users interacting with an agent. Koda and Maes (1996) found that the physical appearance of an agent only affected people's perceptions of intelligence if they did not interact with the agent. If there was interaction involved then the agent (in this case an opponent playing poker) was perceived to be equally intelligent regardless of whether it was visualised by a caricature face or not visualised at all. This suggests that the physical appearance did not influence judgements of perceived intelligence, and that attractiveness and intelligence may be two separable attributes. They also found, however, that participants preferred a poker game when playing with an opponent who was visualised rather than an invisible opponent. In contrast to this, Sproull et al. (1996) found that participants rated a virtual counsellor as more attractive and friendly when they were presented with written text rather than an animated face. Dehn and van Mulken (2000) explained this inconsistency in the perception of agents as being attributable to the type of agent chosen. This again supports the idea that the type of agent and its physical appearance are important factors when choosing agents to be used in an interface.

According to the multiple linear regression analysis, beauty did not have a significant impact on the perceived usefulness of an agent. However, intelligence, overall impression and appropriateness of the agent were the best predictors of how useful participants thought the agent would be. This suggests that the 'what is beautiful is good' hypothesis may not always be true, or may be true only up to certain limits or in certain circumstances, and that an attractive agent may not be the best choice for use on a financial company's website, where factors such as perceived intelligence may be of greater importance. However, the relationship between beauty and utility may be curvi-linear, and, if so, would not be revealed in the current analysis. What may be required is to find the balance, as opposed to finding the most attractive agent, in order for it to be thought of as attractive, intelligent and useful.

4.5 Limitations of Approach

However, too much must not be extrapolated from the data presented here. Perhaps the most obvious limitation to the present study was the range of agents used. The sample of agents was selected on the basis of category membership from a population of around 150 (currently in use/or that could be used on the web). The categories were 'human looking', 'cartoon humans' (caricatures), 'cartoon animals', 'cartoon objects', and 'cartoon human looking'. The results showed that the female agents were preferred, regardless of which category they fell into. This may suggest that the sample of agents used was biased to include more attractive female agents. Similarly, the male agents were liked least, which again could suggest that the sample was biased with a high proportion of unattractive male agents. An alternative explanation could be that there is a culturally agreed standard of female attractiveness, whereas there does not appear to be the same standard for males.

An additional limitation was the extent to which participants envisaged the agents in context. For some participants, the context in which they were told the agent would be used may have been more of a salient influence than for other participants. As there is no way of knowing this, it is impossible to determine whether participant's perceptions of the agents were based solely on physical appearance, or were based on how the agents would be perceived in a financial provisions setting.

Another factor that may have biased judgements was that Task 1 was always completed before Task 2, which may have caused task related order effects. This may have influenced participants to make judgements about what they thought of the agent, perhaps even categorising them, while doing Task 1. The effect of this would be that the judgements made in Task 2 were not a true reflection on the *initial impression* of the agent. Another criticism is that the category scales used in Task 2 may not have been the most salient or most appropriate. As Kelly (1955) proposed, our theories about what other people are like are unique, and it may be useful in future studies to consider what the most salient features of an agent are to the individual doing the judging.

The Eysenck Personality Inventory (EPI) used in Task 3 is a self-report questionnaire, which many participants found difficult to use when rating the personality of the agents they liked and disliked. Some participants commented that they had difficulty in relating the questions to the agents, thus the scores for the agents may not have been a true reflection of the agents, or may have been inconsistent across participants. A final criticism of the experiment was the amount of time taken to complete it. The overall time taken ranged from 45 minutes to 80 minutes, approximately, so that response fatigue may have affected not only the experiment as a whole, but each or any of the individual tasks.

4.6 Conclusion

Many of the results in this study seem to support the 'what is beautiful is good' theory, in that there appeared to be a general 'halo effect' whereby high ratings on beauty generally meant high ratings on other attributes. The overall impression of an agent was most influenced by aesthetic measures of beauty, appeal and pleasantness, all of which had a positive correlation with other attributes such as intelligence and how useful participants thought the agent would be.

However, there is no a priori reason for selecting attributes that would be perceived as important in an agent associated with financial service provision and a separate study is required to determine if the perceptions of agent attributes match with the critical features that are important in customers' perceptions of such provision. In this study it was initially assumed that intelligence and other utilitarian characteristics would be important. The analysis of the potential impact of attractiveness on these variables will be investigated in this work.

The above studies confirm that there is a strong relationship between aesthetics and the apparent or perceived usability of an interface. This may be significant in the current research when participants are asked to rate agents in the context of a financial company's website. Over-riding factors about the website's aesthetics may influence judgements about how useful the agent is, and measures must be taken to ensure these factors are controlled for. However, in the studies carried out by Tractinsky et al.

(2000), it is not clear what actually made an aesthetically pleasing interface, thus measures must be taken to identify the features that contribute this.

Chapter 5 : Agent and website choices

5.0 Aims of Chapter

In making judgements about agents, a number of confounds have been identified. These include the influence of context, social knowledge, gender and/or gender stereotyping, and the selection of agents used. These factors may clearly have impacted the results found in Chapter 4. Therefore, the aim of this Chapter is to minimize the effect of these factors in order to ascertain the true influence of aesthetics.

5.1 Introduction

One of the main criticisms with the study in Chapter 4 was that judgements made might have reflected the influence of context and not the agent's intrinsic characteristics. It was decided that it was necessary to take out the factor of implied context and to have participants rate a set of agents without context. This would also allow us to select a group of agents that participants considered to be attractive, a group that were unattractive and an average group, irrespective of the context in which they are shown (or imagined).

However, another possible confound in the previous experiment may have been the use of social knowledge and how it may have influenced judgements made. For example, it was found that female agents were preferred to any other group of agents and they were rated highly along a number of dimensions. One possible explanation for this is that many service professions use women as service providers and that women dominate occupations such as shop assistants, bank tellers, and various other service points. Thus, by presenting the items for judgement in the previous experiment within the financial services context it may have been the case that the participants drew upon the knowledge of real world experience, accepting the female agents as more 'usual' or appropriate and therefore viewing them in a more positive light as a consequence. On the other hand, it may simply be that the female agents were inherently more attractive than their male counterparts.

One possible source of distortion or confound could be that many of the graphics used on the Internet are intrinsically driven by the dominance of the male presence on the Internet or male involvement in the Internet industry (such as web designers). Arguably then, the agents constructed for the Internet could be largely satisfying the intentions of the developers or the user groups, dominated by males. However, the most recent surveys have shown that males and females have virtually identical rates of Internet use – around 53.9% for males and 53.8% for females in the United States (NTIA and ESA, U.S. Department of Commerce, 2005), and 53% for males and 47% for females in the UK (e-MORI Technology Tracker, Dec. 2005). While that may be the case, however, the same gender balance is not represented in web design staff, with an estimated 97% of web designers in the UK, for instance, being male (Hamilton, 2002).

More simply, the explanation may lie in the fact that femininity is associated with pleasant characteristics that are pro-social or rewarding in other ways, such as compassion, sympathy and affection (Martin, 1987), and as such female agent presence may be seen as more comforting, supportive or re-assuring for web users. Of course, it is possible that a number of factors may simultaneously influence the judgements made with regard to the agents and therefore a variety of experiments are needed to tease apart these different issues. The first issue to address is the robustness of the current findings with regard to preferences for female agents.

Another issue of concern in the previous experiment was the selection and characteristics of the population of agents tested; accordingly, after a search of the Internet, additional agents were selected to increase the variation in agent qualities. For example, some attempt was made to increase the range of attractiveness of female agents who had been consistently selected as the preferred agent type in the previous study and who had been rated more highly across a range of dimensions other than aesthetic, by both male and female judges. An effort was made to identify female agents who appeared, at face value, to have potentially lower levels of attractiveness. In addition, additional non-human (or cartoon) agents were collected to explore the potential for non-human agents to be used in web agent design as the previous study had indicated that non-human agents were less preferred than female agents - a result that may again have been attributable to the population of agents used. The

construction of agents would have simplified the current selection process of course, but the aim was to investigate the current range of agents in use on the web.

The idea of cognitive disregard - eliminating people who we deem to be unsuitable as friends or acquaintances (Baron & Byrne, 1994) – may also have been an influencing factor. If an agent was deemed unsuitable for the context that was suggested or participants simply do not like the way it looked then it may have been judged more harshly by participants. One of the greatest influences in cognitive disregard is physical appearance, and if an agent was thought to be too young or too old, or too thin or too fat, then this could have greatly influenced opinions of them.

5.2 Phase 1

In the current study the protocol used was similar to that developed by Dion Berscheid & Walster (1972), in which pre-sampling of the target stimuli was employed to identify the different categories of target, as *attractive*, *unattractive* or *average*. Dion and colleagues were investigating whether physically attractive people were thought to be more socially desirable than physically unattractive people. In this study a selection of agents was shown to participants who were asked simply to categorise them on the basis of attractiveness. By doing this it was hoped to generate a selection of agents that could be used in future experiments investigating the influence of aesthetics. In addition, it would allow for confirmation of the general positive bias towards female agents, as compared to the male and non-human agents categories.

A within subjects design was used in this study as all participants completed the same tasks. To control for order effects, the order in which they made the judgements was randomised. The independent variable was gender of the agent, which had 3 levels – male, female and cartoon. The dependent variable was the scores on ratings of like and attractiveness.

It was hypothesised that there would be differences in attractiveness and likeability perceptions between male, female, and cartoon agents, with females being perceived

more favourably, and that these perceptions would be similar for male and female participants.

5.2.1 Method

5.2.1.1 Participants

83 participants took part in this study, 16 males and 67 females, and were recruited by means of convenience sampling from Abertay University.

5.2.1.2 Materials

42 different agents were used as stimuli, 16 male agents, 15 female agents, and 11 cartoon agents (see Appendix 8). The agents were shown in black and white on paper, with each image being approximately 1" x 1" in size. Under each image were 2 bi-polar scales ranging from 1 to 10; one scale was for ratings of attractiveness, and the other for rating the likeability of the agent shown. There were 6 agents shown on each A4 page.

5.2.1.3 Procedure

Participants were given a booklet containing the 42 agents and asked to rate the attractiveness of the agents on a 1 to 10 scale, where 1 was unattractive and 10 was attractive. To maintain consistency with Experiment 1, a 10-point scale was used in this study. Participants were also asked to rate how much they liked each agent on a similar scale, where 1 was dislike and 10 was like. The order in which the agents appeared was randomised across participants.

5.2.2 Results

In order to investigate any gender differences between participants, two separate Mann Whitney analyses were carried out, one investigating attractiveness score and the other likeability scores. Results showed that there were no significant differences between male and female participant's attractiveness scores ($U=492$, $N1 = 16$, $N2 = 67$, $p = 0.611$), and no significant differences between male and female participant's

likeability scores (U=498, N1 =16, N2 = 67, p = 0.665). Therefore, it was decided to collapse across gender.

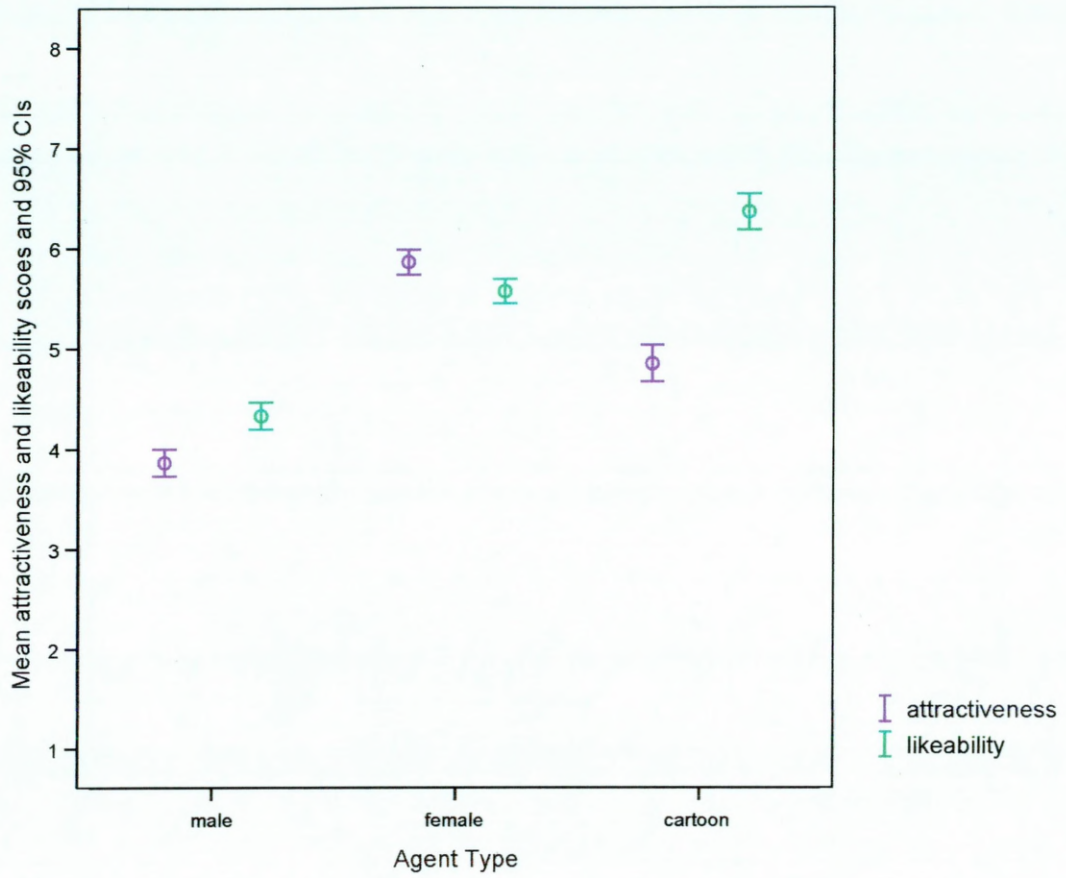


Figure 5.1: Mean Attractiveness and Likeability scores for Male, Female and Cartoon Agents, on a 1-10 scale

Figure 5.1, above, shows that the mean Attractiveness score for male agents was 3.87, with the 95% confidence level falling between 3.74 and 4.00, the mean Attractiveness score for female agents was 5.86, with the 95% confidence level falling between 5.74 and 5.99, and the mean Attractiveness score for cartoon agents was 4.85, with the 95% confidence level falling between 4.67 and 5.03. The mean Likeability score for male agents was 4.33, with the 95% confidence level falling between 4.20 and 4.46, the mean Likeability score for female agents was 5.57, with the 95% confidence level falling between 5.45 and 5.69, and the mean Likeability score for cartoon agents was 6.36, with the 95% confidence level falling between 6.18 and 6.54.

The mean attractiveness scores suggest that females agents are thought to be most attractive, followed by cartoons agents, with male agents thought to be least attractive

(the higher the score the more attractive they are thought to be). However, the mean scores for how much the agents were liked showed that overall, cartoons agents are liked most, followed by females, with males liked the least.

In order to test for a relationship between ratings of attractiveness and liking on each of the three agent categories three Spearman's Rho correlations were carried out, and it was found that statistically significant relationships existed between how attractive and agent is and how much it is liked. However, it should also be noted that the relationship between attractiveness and liking for cartoon agents is more weakly correlated, compared to the correlations for male agents and female agents.

Table 5.1

Correlation between attractiveness and liking.

Agent	Statistical Summary
Male	$r = 0.626^*$, $p<0.0001$
Female	$r = 0.709^*$, $p<0.0001$
Cartoon	$r = 0.421^*$, $p<0.0001$

* Correlation significant at 0.01 level

In order to test for differences between male, female and cartoon agents on ratings of attractiveness, a Mann Whitney U test was carried out and analysis revealed that female agents were perceived as being significantly more attractive than male agents ($U=21$, $n1=16$, $n2=15$, $p<0.001$), and significantly more attractive than cartoon agents ($U=34.5$, $n1=15$, $n2=11$, $p<0.05$). Cartoon agents were considered significantly more attractive than male agents ($U=44$, $n1=16$, $n2=11$, $p<0.05$).

In order to test for differences between male, female and cartoon agents on ratings of likeability, a Mann Whitney U test was carried out and analysis revealed that female agents were liked significantly more than male agents ($U=28.5$, $n1=16$, $n2=15$, $p<0.001$); cartoon agents were also liked more than male agents ($U=2.0$, $n1=16$, $n2=11$, $p<0.001$), but cartoon agents were liked significantly more than female agents ($U=42$, $n1=15$, $n2=11$, $p<0.05$).

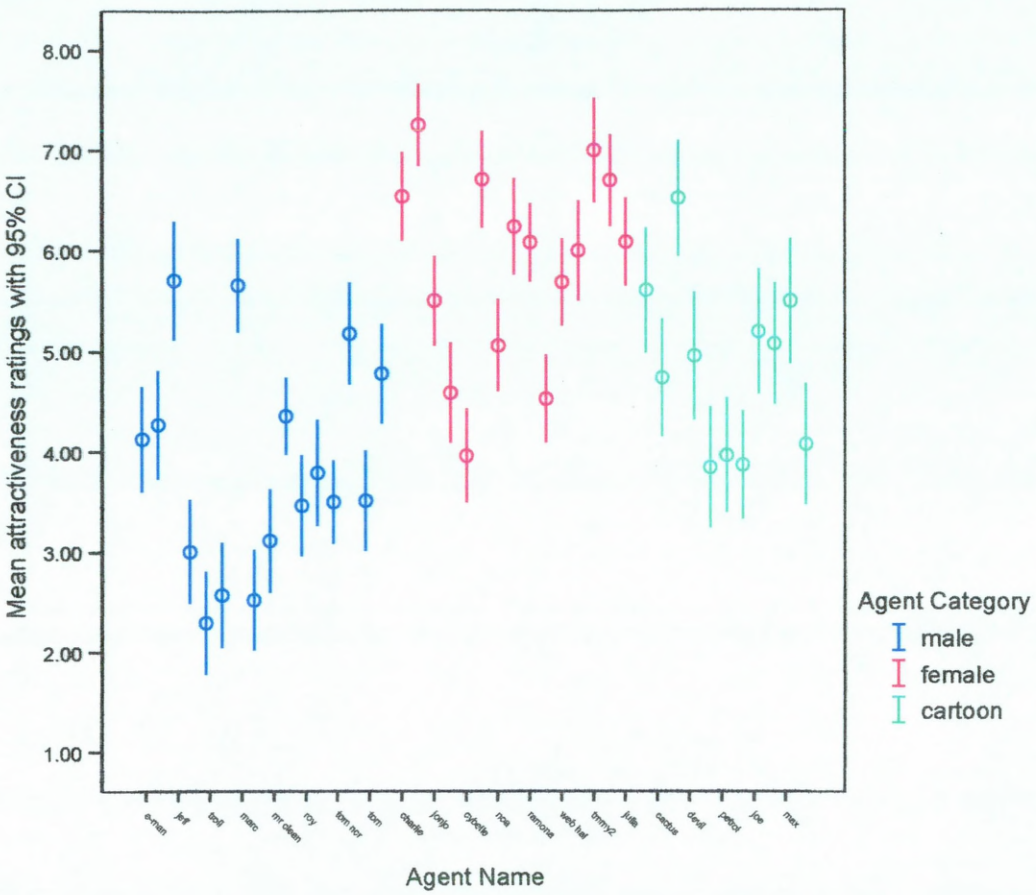


Figure 5.2: Overall Mean Rating of Attractiveness of Individual Agents (Phase1), on a 1-10 scale

From the overall mean ratings of attractiveness (as shown in Figure 5.2), it was decided to use the 2 most attractive agents, 2 least attractive agents and an average agent from each of the three agent categories (female, male and cartoon), as shown in Figure 5.3 below, in future experimentation (Chapter 6). This range of ratings of agents will facilitate the determination of an effect of aesthetics on perceptions of agents.

Attractive Agents	
Tmmy	
Motorola	
Jeff	
Marc	
Bonzi	
Angel	

Average Agents	
Ananova	
Rudiger	
Devil	

Unattractive Agents	
Cybelle	
She-mail	
Troll	
MP	
Ultra-Hal	
Monster	

Figure 5.3: Attractive, Average and Unattractive Agents

5.3 Phase 2

A similar study was conducted using a slightly different set of agents, which tried to include more 'attractive' male agents. To ensure reliability of techniques, a replication study was run Phase 1 (above) involved 6 agents being shown on an A4 page. It was thought that this might influence responses as participants may have compared the six agents shown. Thus, each agent was shown on a separate page (the size of the images remained the same as in Phase 1). As well as rating attractiveness and likeability of the agent, participants were also asked to rate the agent on a further seven dimensions. This would allow the ratings from this study to be used as baseline measures for any future studies involving the agents chosen and scores across studies could be compared to check for consistency. In addition, 26 of the agents were the same as were used in Phase 1, therefore cross-study comparisons of attractiveness and likeability could be made to check if these judgements remained consistent across a different sample of participants.

A within subjects design was used in this study. The independent variable was gender of the agent, which had 3 levels – male, female and cartoon. The dependent variable was the scores on each of the nine attributes.

It was hypothesised that there would be differences in attractiveness perceptions between male, female, and cartoon agents, with females being regarded as more attractive, and that these perceptions would be similar for male and female participants.

5.3.1 Method

5.3.1.1 Participants

16 participants took part in this study, 7 males and 9 females, all of whom were recruited from the Abertay University by means of convenience sampling.

5.3.1.2 Materials

30 different agents were used as stimuli (see Appendix 9) – 10 male agents, 10 female agents, and 10 cartoon agents. The agents were shown in black and white on paper, with each picture being approximately 1" x 1" in size. Under each picture were 9 bi-polar scales ranging from 1 to 10; one scale was used for each of the nine attributes on which the agent was to be judged. These dimensions were attractiveness, trustworthiness, sensibleness, appeal, usefulness, memorability, intelligence, friendliness, and likeability. Each agent and set of rating scales were shown on a separate page, approximately A7 in size.

5.2.1.3 Procedure

Participants were given a booklet containing the 30 agents and were asked to rate each agent on nine dimensions, along a 10-point bi-polar scale. Each agent was shown on a page (approximately A7 size) and the rating scales given below the image. The agents were presented in a random order in each booklet, and there was no time limit on the participants in making their judgements.

5.3.2 Results

In order to test for any significant differences between male and female participants' ratings on the category scales of the ratings of all agents, a Mann-Whitney test was carried out. It was found that there were no significant differences between male and female participants on 7 of the 9 category ratings (the two that were significant were memorability and friendliness), thus it was decided to collapse across gender ratings.

Table 5.2

Mean Score for Male, Female and Cartoon Agents (with standard deviations in parenthesis)

	Agent type					
Attribute	Male		Female		Cartoon	
Attractiveness	4.45	(1.27)	5.96	(0.85)	4.42	(1.39)
Trust	4.75	(0.70)	6.01	(0.66)	4.52	(0.78)
Sensible	5.18	(1.26)	5.98	(1.08)	3.23	(0.57)
Appeal	4.49	(1.01)	5.88	(0.81)	4.83	(1.12)
Usefulness	5.01	(0.62)	5.96	(0.72)	3.93	(0.64)
Memorable	5.16	(0.76)	5.63	(0.34)	6.72	(0.54)
Intelligence	4.94	(0.99)	6.09	(1.06)	3.67	(0.58)
Friendly	5.68	(0.70)	6.38	(0.54)	6.87	(1.41)
Like	4.37	(0.68)	5.76	(0.70)	5.04	(1.11)

Table 5.2 shows that the mean attractiveness scores suggest that females agents were thought to be most attractive, followed by males, with cartoons thought to be least attractive (the higher the score the more attractive the agents are thought to be). Female agents were also rated more highly on trustworthiness, sensibleness, appeal, usefulness, intelligence, and how much they were liked. Cartoon agents were thought to be more memorable and friendly than either male or female agents.

In order to test these observed differences, a Mann-Whitney analysis was carried out, and the results are shown in tables 5.3, 5.4, and 5.5 below.

Table 5.3

Statistical summaries of male v female agents on nine attributes

Attribute	Direction of result	Statistical Summary[a]
Attractiveness	Female > Male	U =15.0, p = 0.008
Trustworthiness	Female > Male	U = 6.5, p = 0.001
Sensibleness	Not Significant	U = 31.5, p = 0.161
Appeal	Female > Male	U =14.5, p = 0.007
Usefulness	Female > Male	U = 17.0, p = 0.012
Memorability	Female > Male	U = 22.5, p = 0.037
Intelligence	Female > Male	U = 19.0, p = 0.019
Friendliness	Female > Male	U = 22.5, p = 0.037
Likeability	Female > Male	U = 8.5, p = 0.002

[a] n1 = 10, n2 = 10

Table 5.4

Statistical summaries of male v cartoon agents on nine attributes

Attribute	**Direction of result**	**Statistical Summary[a]**
Attractiveness	Not significant	U = 45.0, p > 0.05
Trustworthiness	Not significant	U = 41.0, p > 0.05
Sensibleness	Male > Cartoon	U = 9.0, p = 0.002
Appeal	Not significant	U = 43, p > 0.05
Usefulness	Male > Cartoon	U = 11, p = 0.003
Memorability	Cartoon >Male	U = 6.0, p = 0.001
Intelligence	Male > Cartoon	U = 13.0, p = 0.005
Friendliness	Cartoon >Male	U = 20.0, p = 0.023
Likeability	Not Significant	U = 29.0, p > 0.05

[a] n1 = 10, n2 = 10

Table 5.5

Statistical summaries of female v cartoon agents on nine attributes

Attribute	Direction of result	Statistical Summary
Attractiveness	Female > Cartoon	U = 18.0, p = 0.015
Trustworthiness	Female > Cartoon	U = 55.0, p = 0.001
Sensibleness	Female > Cartoon	U = 2.0, p = 0.006
Appeal	Female > Cartoon	U = 20.5, p = 0.026
Usefulness	Female > Cartoon	U = 2.0, p <0.0001
Memorability	Cartoon > Female	U = 4.0, p = 0.001
Intelligence	Female > Cartoon	U = 4.0, p = 0.001
Friendliness	Not Significant	U = 27.5, p >0.05
Likeability	Not Significant	U = 26.5, p >0.05

[a] n1 = 10, n2 = 10

Results showed that there were significant differences between male and female agents on eight out of the nine attributes (the only exception was sensibleness), with female agents being rated more favourably on each of them. There were also significant differences between female and cartoon agents on seven out of the nine attributes, with female agents being rated more favourably on six out of nine (the only exceptions were friendliness and likeability). The only attribute on which cartoons were judged more favourably than female agents was memorability. Similarly, cartoon agents were rated more favourably on memorability and friendliness than male agents. However, male agents were rated as being significantly more sensible, useful and intelligent than cartoon agents. Specifically on ratings of attractiveness, female agents were considered to be significantly more attractive when compared to male agents and cartoon agents. However there was no difference between male agents and cartoon agents on ratings of attractiveness.

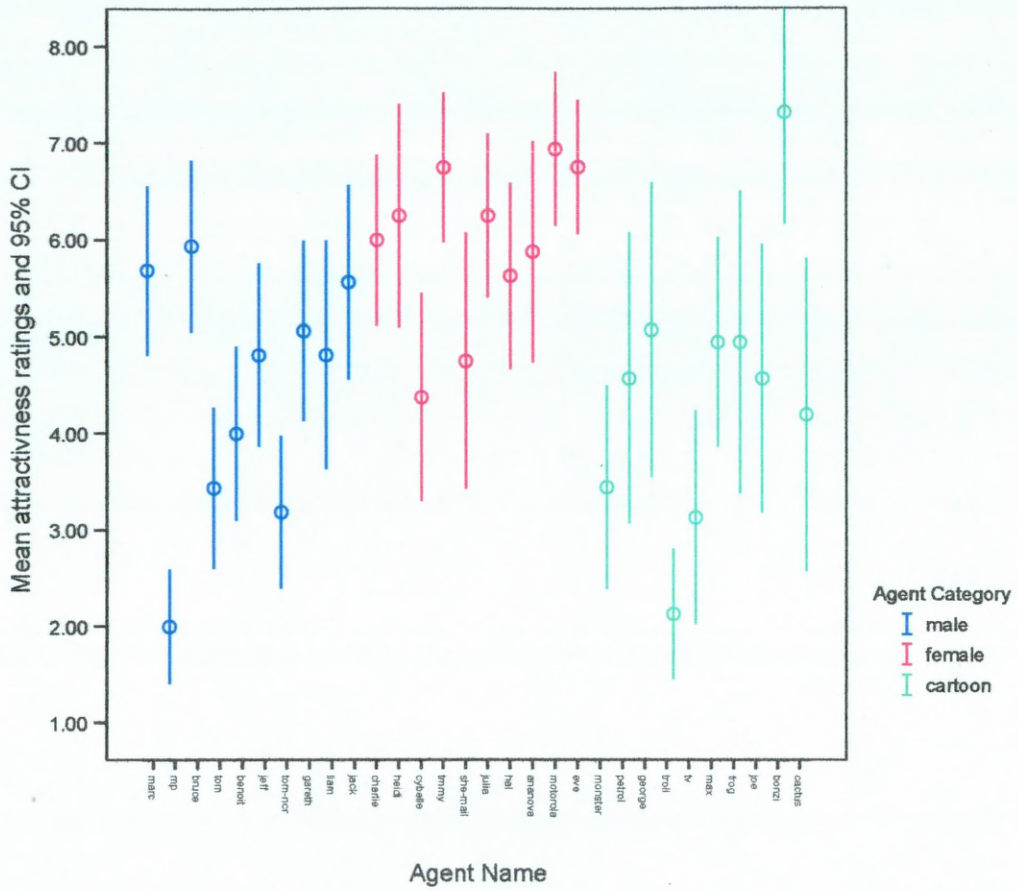


Figure 5.4: Overall Mean Rating of Attractiveness of Agents (Phase2), on a 1-10 scale.

From the overall mean ratings of attractiveness (as shown in Figure 5.4), it was decided to use 2 male and 2 female attractive agents and an attractive cartoon agent, 2 male and 2 female unattractive agents and an unattractive cartoon agent, and 2 male and 2 female average agents and an average cartoon agent, as shown below. In doing this we may be able to determine in future experiments whether there is an effect of aesthetics. As the implied context for the agents in Chapter 4 was a financial company, and because this context would be used in future studies, it was envisaged that cartoon agents would be considered less 'appropriate', therefore it was decided to only include one cartoon agent in each of the attractiveness categories in order to investigate simply the role of aesthetics of cartoon agents in future studies.

5.4 Phase 1 and Phase 2 comparisons

Tables 5.6, 5.7, and 5.8, below, show the mean ratings on each of the nine attributes from Phase 1 and Phase 2, by attractiveness groupings.

Table 5.6

Attractive Agents with overall mean attractiveness ratings in Phase 1 and Phase 2

Agent Name	Agent	Phase 1	Phase 2
Tmmy		7.000 (2.3530)	6.7500 (1.4376)
Bonzi		6.5181 (2.5916)	7.3125 (2.1515)
Motorola		6.7108 (2.1726)	6.9375 (1.4818)
Marc		5.6627 (2.0911)	5.6875 (1.6215)
Bruce		—	5.9375 (1.6520)

Table 5.7

Average Agents with overall mean attractiveness ratings in Phase 1 and Phase 2

Agent Name	Agent	Phase 1	Phase 2
Jeff		5.7108 (2.6528)	4.8125 (1.7595)
Ananova		6.000 (2.2469)	5.8750 (2.1252)
Charlie		6.5422 (1.9773)	6.000 (1.6333)
Benoit		4.7831 (2.2309)	4.000 (1.6733)
Cactus		4.7349 (2.6276)	4.1875 (3.0159)

Table 5.8

Unattractive Agents with overall mean attractiveness ratings in Phase 1 and Phase 2

Agent Name	Agent	Phase 1	Phase 2
She-mail		4.5301 (1.9652)	4.750 (2.4631)
MP		2.5301 (2.2488)	2.000 (1.0954)
Troll		2.3012 (2.3150)	2.1250 (1.2583)
Tom		3.5060 (1.8700)	3.1875 (1.9705)
Cybelle		3.9639 (2.0921)	4.3750 (1.9958)

In order to test for any differences between Phase 1 and Phase 2, a Mann Whitney test was carried out for each attractiveness category. Results showed that there was no significant difference between Phase 1 and Phase 2 in the attractiveness category ($U=5.0$, $n1=4$, $n2=4$, $p>0.05$), nor in the average category ($U=8.5$, $n1=5$, $n2=5$, $p>0.05$), or the unattractive category ($U=11.0$, $n1=5$, $n2=5$, $p>0.05$). In addition, separate Mann-Whitney analyses were carried out on each individual agent and results showed no significant differences between Phase 1 and Phase 2 for any of the agents.

This suggests that the attractiveness ratings remain stable across different populations. With this set of 15 agents (5 attractive, 5 of average attractiveness, and 5 unattractive), the role of physical attractiveness can be further investigated in agent design.

5.5 Phase 3

Phase 1 and 2 were concerned with agent ratings, and it was found that rating remained consistent across the phases. However, the other half of the decision space is context. In order to assess the influence of context in future experiments, it was first necessary to obtain a set of websites that could be used. Following on from the service provision context in Chapter 4, it was decided to collect a set of bank/financial websites that could be used. As previous studies have shown (for example, Tractinsky, 1997) the aesthetics of an interface can influence other judgements made. In a similar manner to the agent selection process, it was decided to show a selection of websites to participants and collect ratings of various dimensions of websites, which would allow categorisation of the websites into different aesthetics groupings.

A within subjects design was used in this study. The independent variables were the websites used. The dependent variables were the rating scores on each of the website attributes. It was hypothesised that there would be differences in perceptions of the different websites, but that these perceptions would be similar for male and female participants.

5.5.1 Method

5.5.1.1 Participants

10 participants took part in this study, 5 males and 5 females. They were recruited from Abertay University by means of convenience sampling.

5.5.1.2 Materials & Apparatus

34 different web sites were used as stimuli (see Appendix 10). These were obtained as screenshots from the Internet and were all bank/financial websites, from different countries (e.g. Barclays, Bank of China, and Egg (internet bank)). The study was run on a PC using SuperLab Pro for Windows, version 1.04. The laboratory location, lighting and viewing distance were kept constant for all participants. A questionnaire was administered to rate the web sites on nineteen different dimensions, on a 10-point

bi-polar scale (see Appendix 11). The dimensions were beauty, appropriateness, memorability, warmth, fun, appeal, trustworthiness, friendliness, elegance, interest, structure, usability, logical layout, calmness, intuition, transparency, clutter, professionalism and overall impression. These categories were based loosely on the category scales used by Schenkman and Jonnson (2000) in their study of webpage aesthetics and preferences.

5.5.1.3 Procedure

The 34 websites were shown in a random order, individually on a PC. The task was to judge each of the web sites on a set of nineteen dimensions such as beauty, appropriateness, and overall impression. A set of instruction/answer sheets was given prior to beginning the task. Each website was allocated a number, and participants had to make a note of this number at the top of their ratings questionnaire. The study was self-paced, with participants pressing the 'return' key to move onto the next website.

5.5.2 Results

Means and standard deviations for each of the 34 websites on each of the nineteen dimensions can be found in Appendix 12. The mean scores showed that Barclays was rated highest on eleven out of the nineteen dimensions (these were appropriateness, appeal, trust, elegant, structured, usable, logical intuitive, transparent, professional, and overall impression). Websites that were also rated highly were Sainsburys (warmth, fun, friendliness, and interesting), Maerki Baumann & Co (beauty, calm, and uncluttered), Tesco (memorable), and Bank of Montreal (logical).

A multiple regression was performed (using the Enter method) to determine the best predictors of overall impression of a website. The model predicting overall impression included eighteen predictors (beauty, appropriateness, memorability, warmth, fun, appeal, trustworthiness, friendliness, elegance, interest, structure, usability, logical layout, calmness, intuition, transparency, clutter and professionalism), and accounted for a significant amount of variance [adjusted R^2 = 0.141, $F(18,321) = 3.923$, $p<0.0001$]. However, none of the individual attributes were

significant predictors in this model, and the model only predicted 14% of the variance in scores. In order to further understand how the website attributes related to each other, a principal components factor analysis was conducted on the correlations of the nineteen website attributes. Three factors were initially extracted with eigenvalues equal to or greater than 1.00. Orthogonal rotation of the factors yielded the factor structure given in Table 5.9. The first factor accounted for 27.6% of the variance, the second factor 22.9%, and the third factor 20.6%. The first factor seemed to be related to the layout of the website, the second factor to the website appeal, and the third factor to the website aesthetics. Figure 5.5 shows the factor loadings of the first two factors for the website attributes.

Table 5.9

Orthogonal factor loading matrix for nineteen website attributes

Variable	Factor 1	Factor 2	Factor 3
Structured	**0.771**	0.131	0.346
Usable	**0.813**	0.250	0.279
Logical layout	**0.875**	0.181	0.225
Calm	**0.757**	0.216	0.329
Intuitive	**0.784**	0.319	0.229
Transparent	**0.774**	0.268	0.254
Uncluttered	**0.735**	0.249	0.238
Overall impression	**0.342**	0.168	0.247
Memorable	0.372	**0.760**	0.098
Warm	0.204	**0.887**	0.075
Fun	0.158	**0.903**	-0.093
Appeal	0.321	**0.624**	0.576
Friendly	0.246	**0.852**	0.157
Interesting	0.218	**0.825**	0.365
Beauty	0.237	0.457	**0.705**
Appropriate	0.262	0.007	**0.793**
Trustworthy	0.287	0.098	**0.736**
Elegant	0.365	0.355	**0.694**
Professional	0.352	-0.088	**0.806**

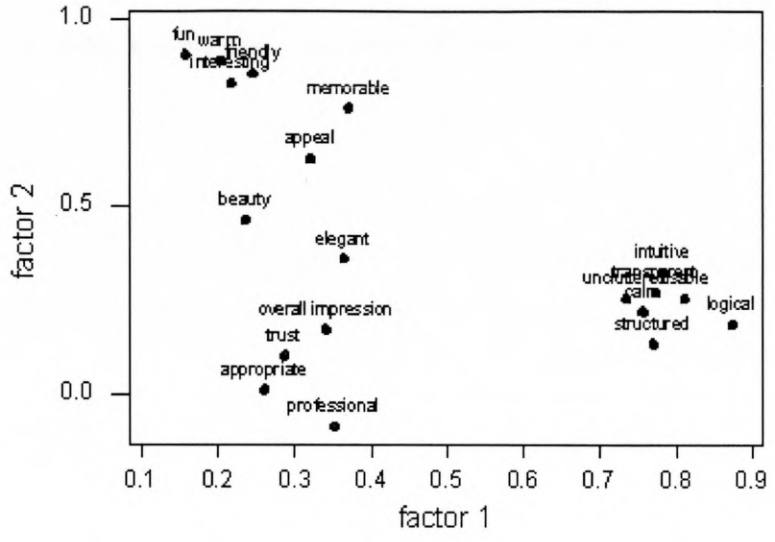


Figure 5.5: Factor loadings for the website attributes of Factor 2 plotted against those of Factor 1

Thus, as there were no specific factors that most influenced the overall impression of a website, and there was no a priori reason for selecting any one attribute to determine what would constitute a good or bad website, it was decided to use the mean rating for all attributes as the basis for website choice. On that basis, it was decided to use 15 websites (as there were 15 agents chosen in the previous study), thus the five websites which were rated highest across all attributes were chosen as 'good' website, the five that were rated lowest were chosen as 'bad' websites, and the five that were closest to the average for each attribute were chosen as 'average' websites (see Figures 5.6 – 5.8, below). In doing this it may be possible to determine in future experiments whether there is an influence of website aesthetics when rating agents embedded in a website.

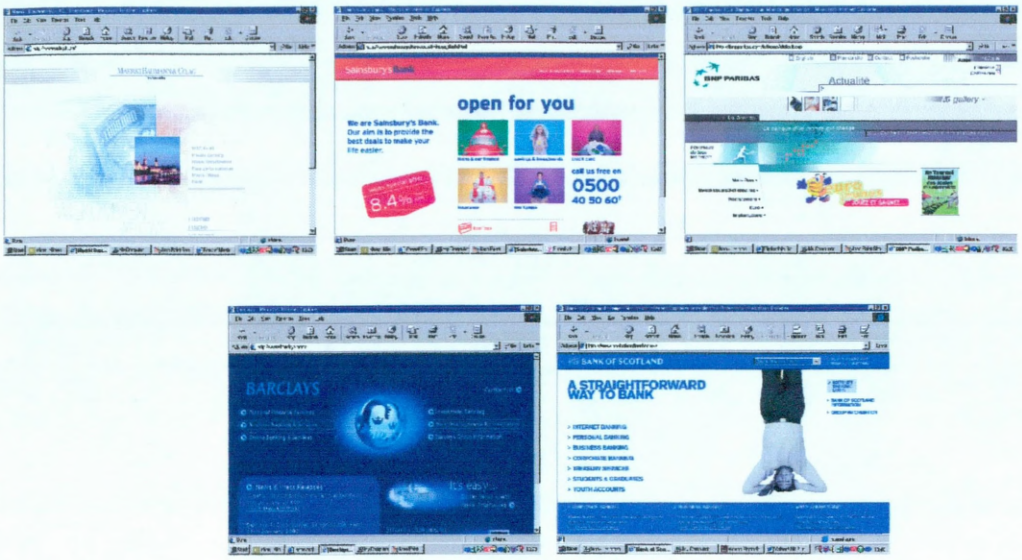


Figure 5.6: "Good" Websites

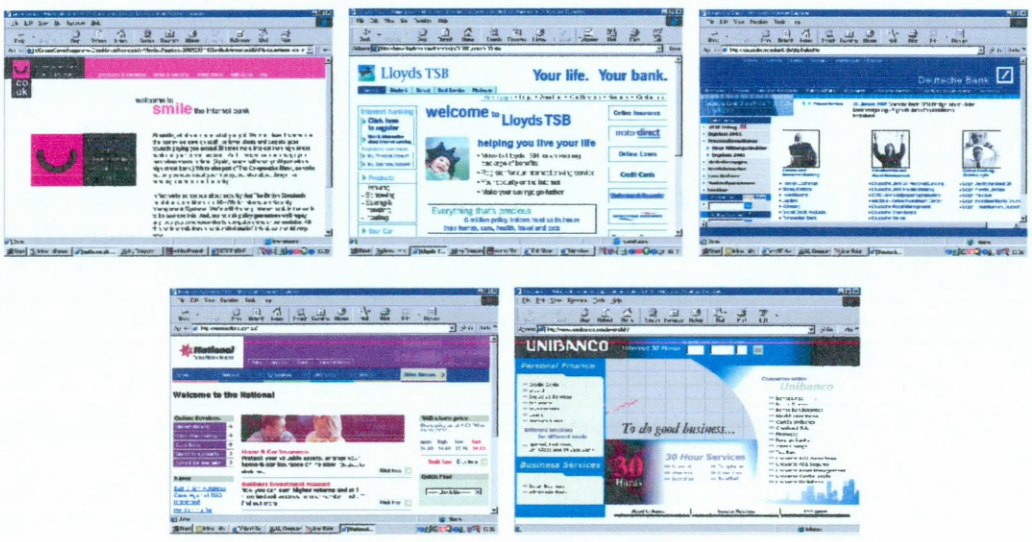


Figure 5.7: "Average" Websites

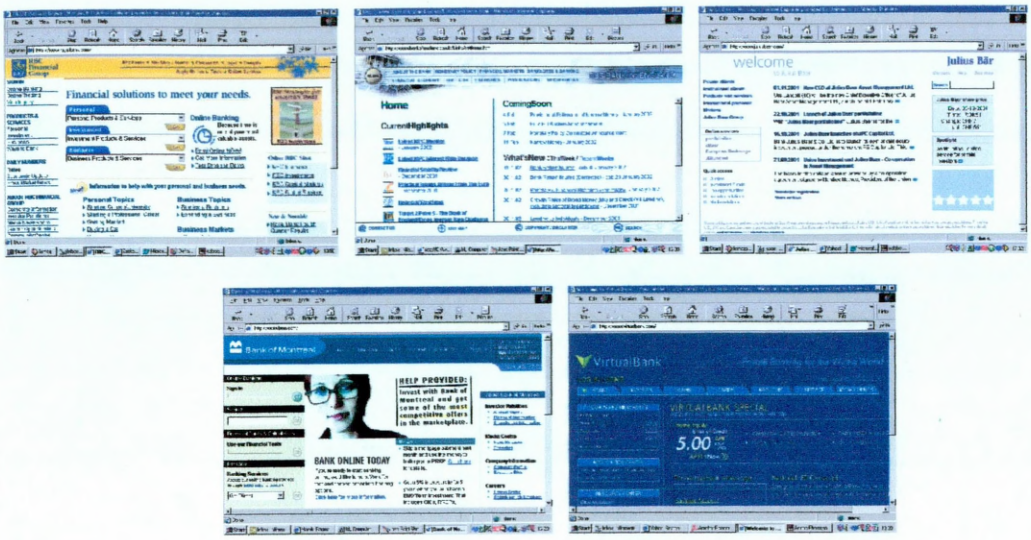


Figure 5.8: "Bad" Websites

5.6 Overall Conclusions

The results from Phase 1 support the hypothesis that there would be an effect of agent type on ratings of attractiveness and likeability. Female agents were rated as being significantly more attractive than both male agents and cartoon agents. In addition, cartoon agents were rated as being significantly more attractive than male agents. However, female agents were not rated as favourably in terms of likeability. Although female agents were liked significantly more than male agents, cartoon agents were liked more than female agents. Although these results suggest that attractiveness may not play a part in ratings of likeability, correlational analysis revealed that there was a significant relationship between ratings of attractiveness and likeability for male, female, and cartoon agents. This relationship was strongest for female agents, followed by male agents, with the relationship for cartoon agents being relatively weak.

The results from Phase 2 replicate the results from Phase 1. Female agents were rated higher compared to male agents and cartoon agents on seven out of the nine attributes measured (attractiveness, trust, sense, appeal, usefulness, intelligence, and likeability). The only two attributes on which female agents were not rated highest for were memorability and friendliness, for which cartoon agents were rated most favourably.

Furthermore, female agents were rated significantly higher on all nine attributes when compared to male agents, and significantly higher on seven attributes when compared to cartoon agents. There were no difference between male agents and cartoon agents on ratings of attractiveness, trust, and appeal, but male agents were considered to be significantly more sensible, useful and intelligent compared to cartoon agents. However, cartoon agents were considered to be significantly more memorable, friendly, and likeable than male agents.

There was no significant difference of ratings of attractiveness between Phase 1 and Phase 2, suggesting that these judgements remain stable.

The results of the current study suggest that female agents are preferred over cartoon agents and that these are in turn generally preferred to male agents. It is not clear if this is attributable to the limited sampling of the potential image range but it does suggest that female agents may, in general, receive more positive responses. While there is some degree of variability in terms of which female agents scored the highest ratings, the segmentation of the agent population into female, male and cartoon agents indicates that certain properties may be more likely to be associated with female agents.

Thus, in future studies the following issues need to be resolved in developing guidelines for web agents. First, whether the properties of female agents are perceived as more desirable than those of male agents and cartoon agents. Second, whether the judgements concerning agent attributes are likely to be influential in the response to web and networked services. The next series of experimental work will examine the interaction between the context of web service provision and agent type.

Chapter 6 : Gender preference and context salience in agent perception

6.0 Aims of chapter

The main aim of this chapter was to establish whether the salience of the context influenced people's perceptions of agents, and to determine whether the preference for female agents was persistent in participants' responses.

In Chapter 4 of this thesis a set of agents were evaluated on a number of dimensions including attractiveness, intelligence and trustworthiness. Participants were told that they were rating agents which were to be used by a bank/financial company on their website to interact with customers and provide assistance. However, there was no confirmatory evidence that participants actually imagined/envisaged the agents being used in this context or that it had any effect on the judgements made. If there was any effect of context then it is impossible to establish whether it was consistent across participants or not, as participant ability to envisage the context may have varied. Thus, there is a likelihood of increased variability along with ambiguity of interpretation of the results. This is important as the effect of the imagined context may have primed certain interpretations of the agents. Thus, some of the judgements made may have more accurately reflected the influence of context rather than the agent's intrinsic characteristics. It was therefore decided that it was necessary to investigate the influence of implied context (which may or may not have been an influencing factor in participants' judgements of the agent) by having participants rate a set of agents with and without context.

6.1 Background

One of the dimensions considered important in product design is appropriateness to the setting in which it will be used. Products that are appropriate, or in fact preferred, in one setting may be completely inappropriate or disliked in another. For example, clothes that are appropriate for a night-club will be inappropriate for a business

meeting. Industrial-style designs such as washing machines found in laundrettes would not be suitable for the home kitchen.

Attractiveness itself may be appropriate or inappropriate to a product. Highly attractive, popular people, mainly women, often advertise products. It reportedly costs advertisers between $1 and $2 million to add this 'dazzle' to their products. In many studies concerning attractiveness, attractiveness is most often defined as, or assumed to be, physical attractiveness, and is simply measured on a single attractive/unattractive dimension (Langmeyer & Shank, 1994).

However, studies such Solomon, Ashmore and Longo's (1992) found that there were six types of physical attractiveness when rating professional female fashion models. These dimensions were classic beauty/feminine, sensual/exotic, sex-kitten, trendy, cute, and girl-next-door. Solomon and his colleagues were interested in whether these specific types of beauty were more or less suitable with certain products (woman's magazines and perfume) when used in advertising. They found that certain beauty types were associated with certain types of magazines and perfumes. For example, Cosmopolitan was associated with the sex-kitten look, and Chanel No 5 was strongly linked with the classic beauty/feminine look. This suggests that advertisers must not only consider how attractive a model is when promoting their product; they must also take into consideration how appropriate they are to that particular product. It is also important to take into consideration the context in which an advertisement will be shown.

It is suggested by Solomon and colleagues that this association can be used to 'convey intended brand images' (p.33), which can, in turn, be vital in advertising campaigns which are mainly image-orientated. These findings could have implications in the current research. If a bank/financial company is trying to portray a particular image then the 'look' of the agent that is chosen for their website could be influential. While the dimensions reported by Solomon et al. (1992) are not being measured in the sample of agents used in the current research it is important to take into account the association between beauty type and product type/image.

Nevertheless, the study conducted by Solomon et al. still only considered different types of physical beauty – a study by Langmeyer and Shank (1994) attempted to

consider all types of beauty and investigate the whole range of characteristics that encompass the concept of beauty. Participants were interviewed about their thoughts on what human beauty was and what product beauty was. The main conclusion was that beauty could be "physical and non-physical; outward and visible, and inward and non-visible" (p.30). Physical beauty was predominantly described as being what can be seen externally, and included facial features, body proportions, healthy appearance, posture, and cleanliness. Non-physical beauty was described as being more about the soul of a person and how spiritual they were. Other traits to describe non-physical beauty included intelligence, liveliness, strength, nurture, innocence, beautiful personality and tolerance. Beautiful products were characterised by the way they worked, how functional they were, and if they were well designed. Design, proportion and shape also defined how beautiful a product was. The main conclusion drawn from this study was that although initial impressions of things (mainly people) are based on physical appearance, a truly beautiful person must also be beautiful inside. "Values, habits, personality, and behaviour are the 'soul' of beauty" (p.34)

The suggestion that intelligence is considered to be a non-physical dimension of beauty supports the results found in the first study, which found that although attractiveness and intelligence were considered to be separate dimensions from each other, there was some degree of correlation between them. The findings in the study by Langmeyer and Shank suggest that perhaps it may be worth considering more what beauty actually means to people and whether the different approaches to beauty can in fact be applied to agents. Arguably, it may be the case that because agents are not 'real people' then they may be viewed as being more like products. In other words, it may be more important for agents to be well designed and functional, rather look attractive and intelligent.

6.1.1 First Impressions and Context

People often try to figure out what others are like by considering how they look on the outside. These perceptions of others often shape the actions towards them, and people treat others as they perceive them and not as they really are. Appearance cues, however, are not very stable and can often lose their salience and power over time and

number of interactions (Jones, 1990). This suggests that while first impressions can greatly influence the views of people, these may not be long lasting. If people react similarly to agents, then over time and interaction the power of the agent may be lost and people will cease to interact with it unless the interaction itself is pleasurable.

6.1.2 Stereotypes

It has been suggested that the effects of the "beauty is good" stereotype are strongest when making judgements about social competence and interpersonal ease (Dion (1981, 1986)), and social vitality or extraversion (Bassili (1981)). The variation in the "beauty is good" stereotype could have major implications when rating or choosing agents to be used on a web interface. If agents are to be chosen based on their physical attractiveness, then this might only be important if the agents have to be perceived to be socially competent. If, however, users think the intelligence of the agent is more important then attractiveness may not be an important salient characteristic. Another possibility could be that if users are interacting with the agent on a regular basis then they may want to feel that it is concerned about them and making decisions for their benefit. Again, attractiveness may not be an important factor if this is the case. These points are particularly relevant because the agents considered in the current research are to be embedded in financial websites, where the importance of factors such as intelligence and trustworthiness may over-ride factors such as attractiveness.

The presence or absence of individuating information has been found to be a significant variable in determining the attractiveness effect (Eagly et al., 1991); hence appearance may only have a weak effect on perceived personal characteristics in a natural setting. This suggests that the website in which an agent is placed, or used, could influence judgements of the agent. Faces often appear in magazines or advertisement which may themselves be positively or negatively evaluated. If the judgement of a person's character is influenced by the context in which a person, or in this experiment an agent, appears then the effect of attractiveness per se may be reduced. However, by contrast, in a study investigating the effect of context on judgements of people, Larose, Tracy and McKelvie (1992) found that the

favourability of the magazine in which the photos were embedded did not influence judgements of the target persons. On the basis of this study it can be suggested that perception of how "good" or "bad" a website is will have no effect on perceptions of the agents.

The current study therefore aimed to assess the influence of context on judgements of agents. It was also hoped that by asking participants whether an agent is suitable for a bank/financial website, it could be determined which attributes 'suitable' agents possess, and therefore what people are looking for in a financial agent.

A between subjects design was employed in this study. The independent variables were the level of context, which had 3 levels – context, imagined context, and no context; the attractiveness of the agents, which had 3 levels – attractive, of average attractiveness, and unattractive; the 'gender' of the agents, which had 3 levels – male, female and cartoon; and the type of website, which had 3 levels – good, average, and bad. The dependent variables were the rating scores on each of the nine agent attributes, the perception of "appropriateness" of the agent to a bank website, and the rating scores on the seventeen website attributes ("context" group only).

It was hypothesised that there would be an effect of context on ratings of agent attributes; an effect of agent attractiveness on ratings of agent attributes; and an effect of agent 'gender' on ratings of agent attributes.

6.2 Method

6.2.1 Participants

53 participants took part in this study - 17 males and 36 females - and were randomly assigned to one of three groups. The first group, Group 1, consisted of 22 participants, 8 males and 14 females. The second group, Group 2, consisted of 15 participants, 4 males and 11 females, and the third group, Group 3, consisted of 16

participants, 5 males and 11 females. All were selected by means of convenience sampling from Abertay University.

6.2.2 Materials & Apparatus

15 different agents were used as stimuli (see Appendix 13), along with 15 different bank/financial websites (see Appendix 14). The agents were categorised into three groups –attractive, of average attractiveness, and unattractive. Similarly, the website were categorised into good, average and bad groups (see Phase 3 in Chapter 5 for further details). The websites and agents were matched up into three separate groups (A, B and C) in order to counterbalance the agent/website combinations. That is, group 'A' consisted of attractive agents/good website, average agents/average websites, and unattractive agents/bad websites; group 'B' consisted of attractive agents/average websites, average agents/bad websites, and unattractive agents/good websites; and group 'C' consisted of attractive agents/bad websites, average agents/good websites, and unattractive agents/average websites.

The trials were run on a PC using SuperLab Pro for Windows, version 1.04. A ratings questionnaire was used by all participants, which required them to mark on a 7-point scale their ratings of a set of agent attributes (these were attractiveness, memorability, appeal, trustworthiness, friendliness, sense, usefulness intelligence, and likeability) (see Appendix 15). Group 1 also had a ratings questionnaire for each of the websites (on which they had to rate each website on appropriateness, memorability, fun, appeal, trustworthiness, friendliness, elegance, interest, structure, usability, logical layout, calmness, intuitiveness, clutter, professionalism, and overall impression) (see Appendix 16).

6.2.3 Procedure

All three groups had to judge each of the 15 agents on 9 attributes, along a 7-point bi-polar scale. Instructions were given verbally prior to beginning the task. The agents were presented, in a random order, individually on a PC with a number preceding

each agent, which participants had to note on the top of each ratings questionnaire. There was no time limit on the participants to make their judgements, and the judgements were made on a separate questionnaire for each agent.

Group 1 (context condition) were told that they were rating agents used by a financial company on their website, and were shown a static image of each website, embedded with one of the 15 agents. They were also asked whether they thought each agent was suitable for the website.

Group 2 (imagined condition) were told that they were rating agents to be used by a financial company on their website but were only a shown image of each agent. After rating the agent participants were asked whether they thought the agent would be suitable to be used by a bank/financial company on their website.

Group 3 (no context condition) were shown a static image of each agent without context, and with no additional information.

Group 1 had the additional task of rating each website on 16 dimensions, including appeal, usability, and overall impression. Again, this was done on a 7-point bi-polar scale, and on a separate questionnaire for each website.

6.3 Results

Within the set of 15 bank websites used in the context condition, 5 of these were considered to be 'good' websites, 5 were 'average' websites, and 5 were 'bad' websites. The websites in each of these categories were chosen on the basis of the results of Study 2. There were also 3 levels of agent attractiveness – attractive, of average attractiveness, and unattractive. Initial analysis was performed in order to investigate the influence the effect of differing levels context (i.e. website levels) on each of the agent attractiveness categories.

The mean scores for each of the agent attributes across each of the agent levels (attractive, average and unattractive) and each of the website levels (good, average and bad) were calculated and can be found in Appendix 17.

Mann Whitney analysis revealed that the only significant differences between agents in each of the context conditions were on the dimensions of memorability [$U = 5088$, $N1 = 120$, $N2 = 105$, $p = 0.011$] and likeability [$U = 5294$, $N1 = 120$, $N2 = 105$, $p = 0.036$]. On both dimensions, agents that were matched with the corresponding website (i.e. attractive agents with good website, average agents with average websites, and unattractive agents with bad websites) were rated significantly lower than when they were not matched (i.e. the difference between group 'A' and group 'B' on the attributes memorability and likeability was significant).

Therefore, groups 'A', 'B', and 'C' were analysed collectively under the 'Context' condition.

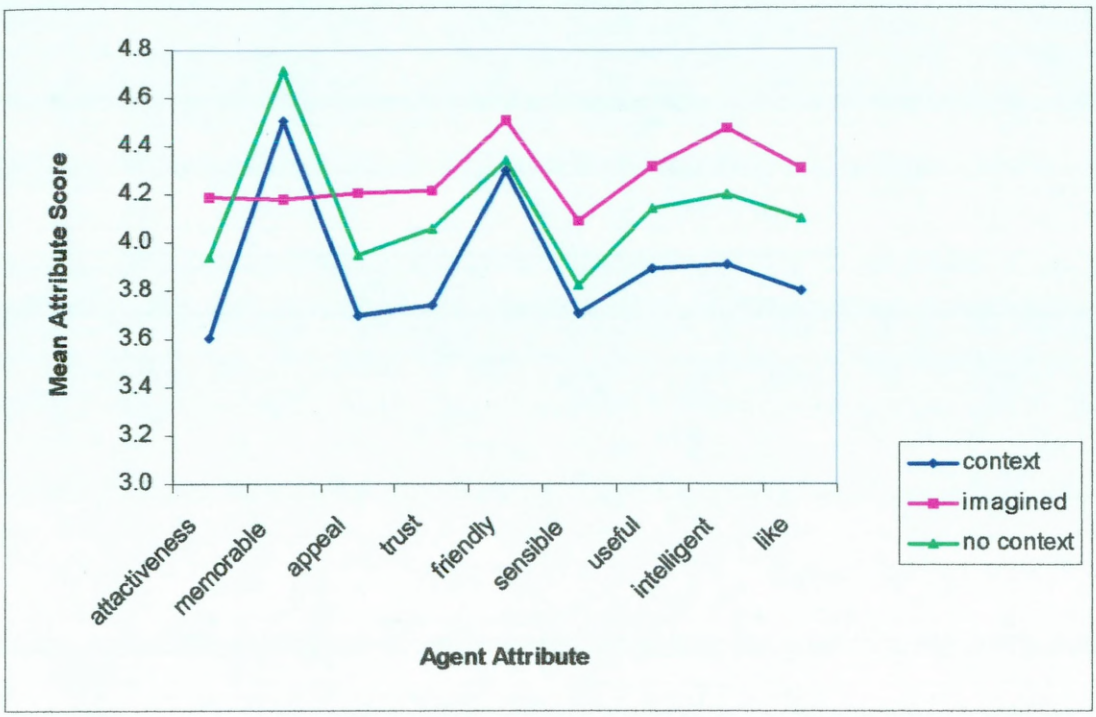


Figure 6.1: Differences between context conditions on each agent attribute

Figure 6.1, above, shows the means ratings on each of the nine agent attributes, under each condition. In general, agents in the context condition were rated lower when compared to the imagined context and no context conditions, on all attributes, with the exception of memorability. Agents in the imagined context condition were rated highest on all attributes, with the exception of memorability.

In order to test the observed differences between conditions on each of the agent attributes, a Mann-Whitney test was carried out for each pair of conditions.

Context v imagined context

It was found that there were significant differences between the context and imagined context conditions on 7 out of the 9 attributes investigated (these were attractiveness, appeal, trustworthiness, sense, usefulness, intelligence, and like). In each case, agents in the context condition were rated significantly lower than in the imagined condition (see Appendix 18 for statistical summary).

Context v no context

There were significant differences between the context and no context conditions on 5 out of the 9 attributes investigated (there were attractiveness, trustworthiness, usefulness, intelligence, and like). In each case, agents in the context condition were rated significantly lower than in the no context condition (see Appendix 18 for statistical summary).

No context v imagined context

The only significant difference between the no context and imagined context conditions was on the attribute memorability, where agents in the imagined context were rated significantly less memorable than in the no context condition (see Appendix 18 for statistical summary).

Agent 'Gender' Results

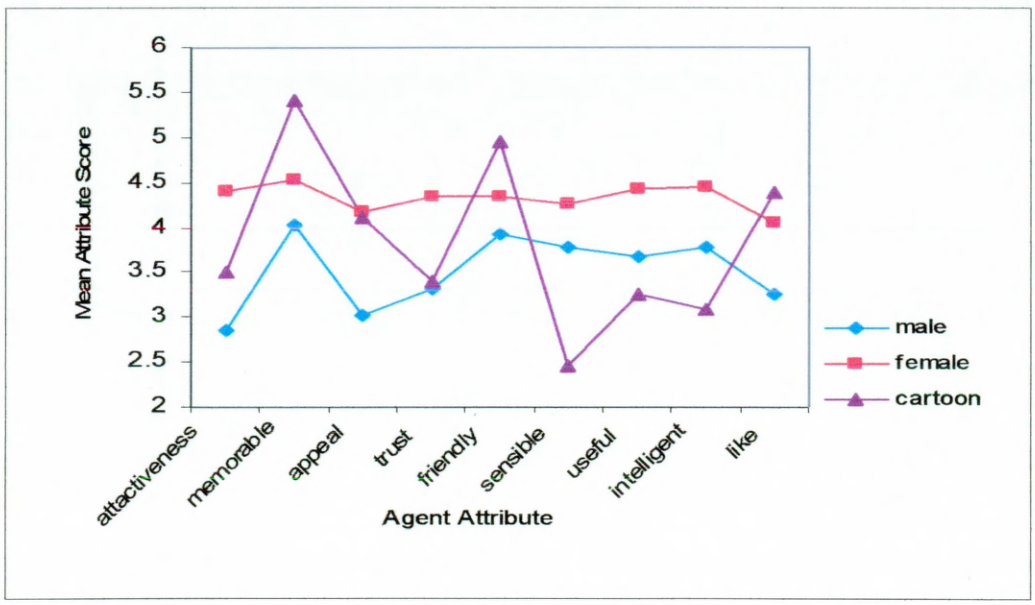


Figure 6.2: Differences between agent 'genders' for each attribute

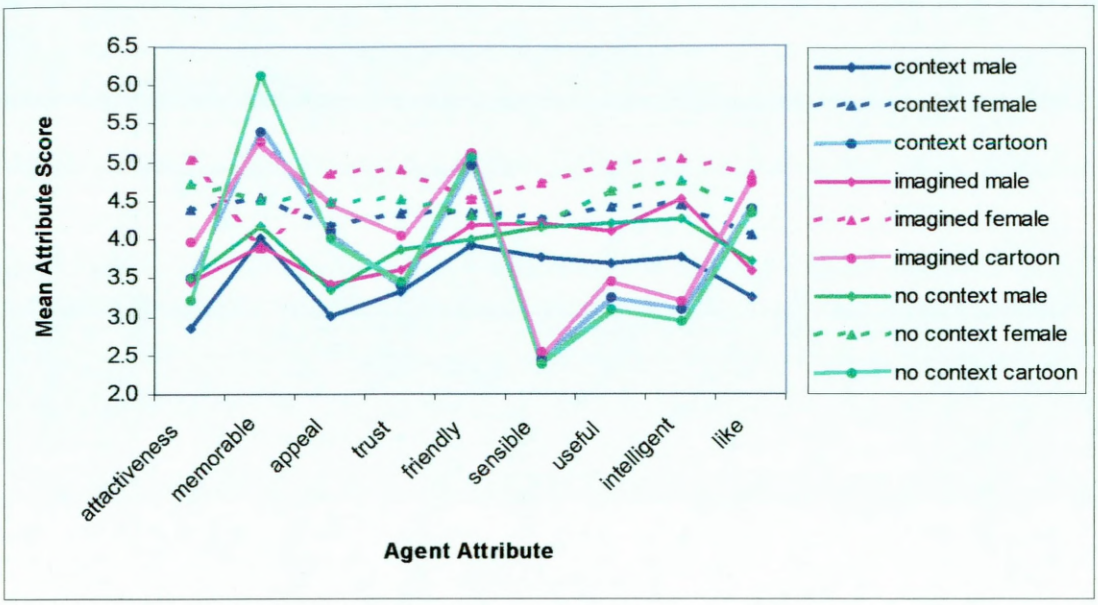


Figure 6.3: Differences between agent 'genders' for each attribute in each context condition

As Figures 6.2 and 6.3 show, there was, over all conditions, a general positive regard for female agents when compared to male agents or cartoon agents. Female agents were considered to be more attractive, trustworthy, sensible, useful and intelligent than male or cartoon agents, and more appealing and likeable than male agents. Cartoon agents were thought to be more memorable and friendly than male or female agents, and more appealing and likeable than male agents. Male agents were thought to be more sensible and intelligent than cartoon agents (see Appendix 19 for table of mean scores).

In order to test the observed differences between conditions on each of the 3 agent types, a Mann-Whitney test was carried out for each pair of conditions

Context v imagined context

There were significant differences between context and imagined context for **female agents** on 8 out of the 9 attributes investigated (these were attractiveness, memorability, appeal, trustworthiness, sense, usefulness, intelligence, and likeability). In each case (except memorability), female agents in the context condition were rated significantly lower than in the imagined condition.

Male agents were judged to be significantly less sensible, useful, and intelligent in the context condition than in the imagined condition.
There were no significant differences between the context and imagined conditions for **cartoon agents** (see Appendix 20 for statistical summary).

Context v no context

There were significant differences between context and no context for **male agents** on 5 out of the 9 attributes investigated (there were attractiveness, trustworthiness, usefulness, intelligence, and like). In each case, male agents in the context condition were rated significantly lower than in the no context condition.
Cartoon agents were judged to be significantly less memorable in the context condition than in the no context condition.
There were no difference between the context and no context conditions for **female agents** (see Appendix 21 for statistical summary).

No context v imagined context

There were significant difference between no context and imagined context for **female agents** on the attributes of attractiveness, memorability, appeal, trustworthiness, sense, usefulness, intelligence and like, with female agents in the no context rated significantly lower on all attributes (with the exception of memorability) than in the imagined context condition.
Cartoon agents were judged to be significantly less memorable in the imagined condition than in the no context condition, but more significantly more attractive in the imagined condition compared to the no context condition.
There were no differences between the imagined and no context conditions for **male agents** (see Appendix 22 for statistical summary).

A summary of the above results is shown in Table 6.1, below. For male and female agents, context has a detrimental effect on ratings when compared to the influence of imagined context. Male agents were also rated significantly lower on over half of the attributes rated when in the context condition compared to the no context condition. However, no context had a detrimental effect on female agent ratings compared to ratings in the imagined context condition.

Table 6.1

Summary of the difference between context conditions for each agent type

Agent type	Context v Imagined Context	Context v No Context	No Context v Imagined Context
Male	3/9 Context ↓	5/9 Context ↓	-
Female	8/9 (7/9 Context ↓)	-	8/9 (7/9 No Context ↓)
Cartoon	-	1/9 Context ↓	2/9

Agent Attractiveness Results

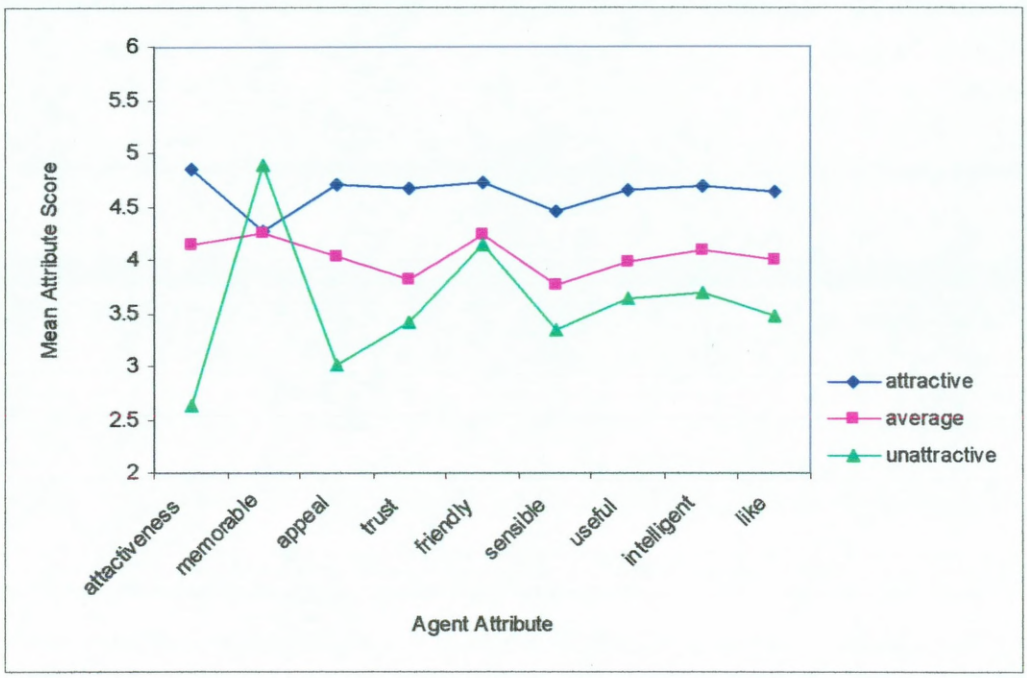


Figure 6.4: Differences between agent attractiveness groups for each attribute

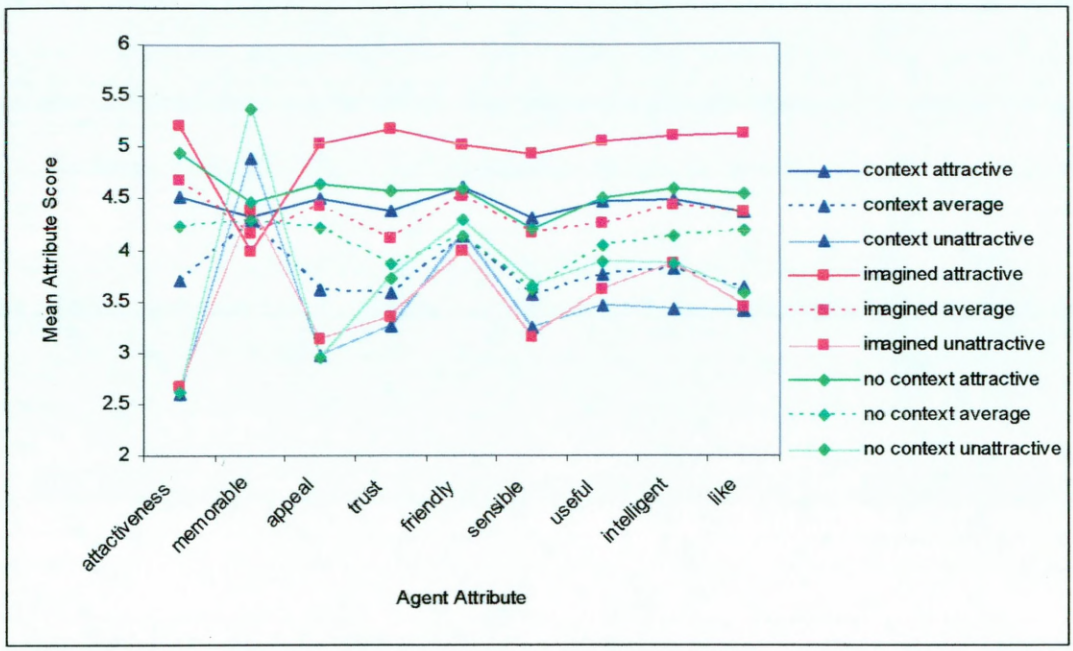


Figure 6.5: Differences between agent attractiveness groups for each attribute in each context condition

As Figures 6.4 and 6.5 shows, there was, over all conditions, a general positive regard for attractive agents when compared to unattractive agents or agents of average attractiveness. Mann Whitney analysis revealed that attractive agents were considered to be significantly more attractive, appealing, trustworthy, friendly, sensible, useful, intelligent, and likeable when compared to agents of average attractiveness. Attractive agents were also considered to be significantly more attractive, appealing, trustworthy, friendly, sensible, useful, intelligent, and likeable when compared to unattractive agents. However, unattractive agents were significantly more memorable than attractive agents. Agents of average attractiveness were also considered to be significantly more attractive, appealing, trustworthy, sensible, useful, intelligent, and likeable when compared to unattractive agents. However, unattractive agents were significantly more memorable than agents of average attractiveness (see Appendix 23 for table of mean scores and Mann Whitney statistical summaries).

In order to test the observed differences between conditions on each of the 3 agent attractiveness groups, a Mann-Whitney test was carried out for each pair of conditions.

Context v imagined context

There were significant differences between context and imagined context for **attractive agents** on 7 out of the 9 attributes investigated (these were attractiveness, appeal, trustworthiness, sense, usefulness, intelligence, and likeability). In each case, attractive agents in the context condition were rated significantly lower than in the imagined condition.

There were significant differences between context and imagined context for **agents of average attractiveness** on 7 out of the 9 attributes investigated (these were attractiveness, appeal, trustworthiness, sense, usefulness, intelligence, and likeability). In each case, agents of average attractiveness in the context condition were rated significantly lower than in the imagined condition.

There were no significant differences between the context and imagined conditions for **unattractive agents** (see Appendix 24 for statistical summary).

Context v no context

Attractive agents were judged to be significantly less attractive in the context condition than in the no context condition.

There were significant differences between context and no context for **agents of average attractiveness** on 3 out of the 9 attributes investigated (these were attractiveness, appeal, and likeability). In each case, agents of average attractiveness in the context condition were rated significantly lower than in the no context condition.

Unattractive agents were judged to be significantly less memorable and useful in the context condition than in the no context condition (see Appendix 25 for statistical summary).

No context v imagined context

There were significant differences between no context and imagined context for **attractive agents** on 5 out of the 9 attributes investigated (these were trustworthiness, sense, usefulness, intelligence, and likeability). In each case, attractive agents in the no context condition were rated significantly lower than in the imagined condition.

Agents of average attractiveness were judged to be significantly less attractive and sensible in the no context condition than in the imagined condition.

Unattractive agents were judged to be significantly less memorable and sensible in the imagined condition than in the no context condition (see Appendix 26 for statistical summary).

A summary of the above results is shown in Table 6.2, below. For attractive and average agents, context has a detrimental effect on ratings when compared to the influence of imagined context. Similarly, no context had a detrimental effect of ratings for attractive and average agents when compared to the influence of imagined context. However, imagined context had a detrimental effect on unattractive agent ratings compared to ratings in the no context condition.

Table 6.2

Summary of the difference between context conditions for each agent attractiveness group

Agent type	Context v Imagined Context	Context v No Context	No Context v Imagined Context
Attractive	7/9 Context ↓	1/9 Context ↓	5/9 No Context ↓
Average	7/9 Context ↓	3/9 Context ↓	2/9 No Context ↓
Unattractive	-	2/9 Context ↓	2/9 Imagined ↓

Regression Analysis

In order to determine the factors that influenced how much an agent was liked, a bivariate multiple regression analysis was performed to determine the best predictors of agent likeablity. Based on results shown above, it was decided to run separate multiple regressions for each condition (context, imagined, and no context) in addition to the overall regression.

Overall

Overall, the model predicting Likeability included eight predictors (attractiveness, memorability, appeal, trustworthiness, friendliness, sense, usefulness and intelligence), and accounted for a significant amount of variance [adjusted $R^2 = 0.655$, $F(8, 801) = 192.733$, $p<0.0001$]. The predictors attractiveness, memorability, appeal,

trustworthiness and friendliness all had a significant impact on predicting how much an agent was liked ($p < 0.001$ in all cases). The impact of the predictors sense, usefulness, and intelligence did not reach significance. The most important predictors were appeal (ß=0.270), followed by friendliness (ß=0.260).

Context Condition

In the context condition, the model predicting Likeability included eight predictors (attractiveness, memorability, appeal, trustworthiness, friendliness, sense, usefulness and intelligence), and accounted for a significant amount of variance [adjusted R^2 = 0.619, $F(8, 321) = 67.942$, $p<0.0001$]. The predictors attractiveness, memorability, appeal, trustworthiness and friendliness all had a significant impact on predicting how much an agent was liked ($p < 0.01$ in all cases). The impact of the predictors sense, usefulness, and intelligence did not reach significance. The most important predictors were trustworthiness (ß=0.308), followed by attractiveness (ß=0.220).

Imagined context condition

In the imagined condition, the model predicting Likeability included eight predictors (attractiveness, memorability, appeal, trustworthiness, friendliness, sense, usefulness and intelligence), and accounted for a significant amount of variance [adjusted R^2 = 0.739, $F(8, 231) = 85.622$, $p<0.0001$]. The predictors attractiveness, appeal, trustworthiness and friendliness all had a significant impact on predicting how much an agent was liked ($p < 0.001$ in all cases). The impact of the predictors memorability, sense, usefulness, and intelligence did not reach significance. The most important predictors were appeal (ß=0.319), followed by attractiveness (ß=0.249).

No context condition

In the no context condition, the model predicting Likeability included eight predictors (attractiveness, memorability, appeal, trustworthiness, friendliness, sense, usefulness and intelligence), and accounted for a significant amount of variance [adjusted R^2 = 0.610, $F(8, 231) = 47.720$, $p<0.0001$]. The predictors attractiveness, appeal, trustworthiness and friendliness all had a significant impact on predicting how much an agent was liked ($p < 0.01$ in all cases). The impact of the predictors memorability,

sense, usefulness, and intelligence did not reach significance. The most important predictors were friendliness (ß=0.387), followed by appeal (ß=0.270).

A summary of the four regression analyses can be found in Table 6.3, below.

Table 6.3

Summary of Regression analysis of agent attributes

Attribute	Overall	Context	Imagined Context	No Context
Attractiveness	√	√ **	√ **	√
Memorability	√	√		
Appeal	√ *	√	√ *	√ **
Trust	√	√ *	√	√
Friendliness	√ **	√	√	√ *
Sense				
Usefulness				
Intelligence				
Adj. R^2	0.655	0.619	0.739	0.610

√ Significant predictor * Most important predictor ** Second most important predictor

Website Analysis

A multiple regression was performed to determine the best predictors of the overall impression of a website. The model predicting overall impression included fifteen predictors (appropriateness, memorability, fun, appeal, trustworthiness, friendliness, elegance, interest, structure, usability, logical layout, calmness, intuitiveness, clutter and professionalism), and accounted for a significant amount of variance [adjusted R^2 = 0.757; $F(15, 314) = 69.170$, $p<0.0001$]. The predictors memorability, appeal, friendliness, interesting, usability, uncluttered and professional all had a significant impact on predicting the overall impression of a website. The impact of the predictors appropriateness, fun, trustworthiness, elegance, structured, logical, calmness, and intuitiveness did not reach significance. The most important predictor was professional (ß=0.285), followed by appeal (ß=0.170).

A factor analysis was performed in order to further understand the relationships among the dimensions. A principal components factor analysis was conducted on the correlations of the 15 variables. Three factors were initially extracted with eigenvalues equal to or greater than 1.00. Orthogonal rotation of the factors yielded the factor structure given in Table 6.4. The first factor accounted for 27% of the variation, the second factor 22% and the third factor 22%. The first factor related to the usability of the website, the second factor related to suitability, and the third related to the warmth of the website.

Table 6.4

Orthogonal factor loading matrix for fifteen website dimensions

Variable	Factor 1	Factor 2	Factor 3
Uncluttered	**.854**	-.022	.197
Calmness	**.852**	.219	.093
Logical	**.707**	.410	.187
Usability	**.624**	.378	.245
Intuitiveness	**.614**	.454	.261
Elegance	**.577**	.510	.306
Structured	**.574**	.449	.279
Trustworthiness	.128	**.860**	.107
Appropriateness	.221	**.796**	.077
Professional	.419	**.752**	-.084
Friendliness	.095	.136	**.835**
Interesting	.293	.154	**.806**
Fun	-.045	-.434	**.767**
Memorability	.365	.130	**.710**
Appeal	.416	.389	**.653**

Appropriateness Judgements Analysis

In order to test for differences in ratings of appropriateness between the context and imagined conditions, Chi-Square analysis was carried out and it was found that there were no significant differences between context and imagined conditions for the agents considered appropriate for the websites (X^2 = 10.459, *df* = 14, p = 0.728). Therefore, overall ratings of appropriateness are reported below.

Table 6.5

Agents considered appropriate for use on a bank website (expressed as a percentage)

		Appropriate		
Agent	Gender	Yes	No	Don't know
Motorola	Female*	89.19	10.81	0.00
Tmmy	Female*	81.08	18.92	0.00
Ananova	Female#	51.35	48.65	0.00
Bruce	Male*	43.24	54.05	2.70
Benoit	Male#	40.54	59.46	0.00
Cybelle	Female+	35.14	62.16	2.70
Tom	Male+	29.73	64.86	5.41
Bonzi	Cartoon*	27.03	67.57	5.41
Marc	Male*	27.03	70.27	2.70
Charlie	Female#	21.62	78.38	0.00
Cactus	Cartoon#	16.22	81.08	2.70
Jeff	Male#	13.51	86.49	0.00
She Mail	Female+	13.51	83.78	2.70
Troll	Cartoon+	10.81	89.19	0.00
MP	Male+	2.70	94.59	2.70

*attractive #average +unattractive

Table 6.5 shows that the agents considered to be most appropriate for use on a bank website were Motorola, followed by Tmmy. The agent that was considered to be least appropriate was MP. Overall, female agents were thought to be most appropriate for a bank website, followed by male agents, with cartoon agents thought to be least appropriate (see figure 6.6).

In order to investigate if there were any significant differences between rating on each agent attribute for the two most appropriate agents, a Mann Whitney test was carried out. It was found that there were significant differences between Motorola and Tmmy on six out of the nine attributes. Tmmy was rated as being significantly more attractive [$U = 1104$, $N1 = 53$, $N2 = 53$, $p < 0.05$], memorable [$U = 929.5$, $N1 = 53$, $N2 = 53$, $p < 0.05$], and appealing [$U = 1078$, $N1 = 53$, $N2 = 53$, $p < 0.05$], when compared to Motorola. However, Motorola was rated as being significantly more sensible [$U = 1060.5$, $N1 = 53$, $N2 = 53$, $p < 0.05$], useful [$U = 979$, $N1 = 53$, $N2 = 53$, $p < 0.05$], and intelligent [$U = 980$, $N1 = 53$, $N2 = 53$, $p < 0.05$], when compared to Tmmy.

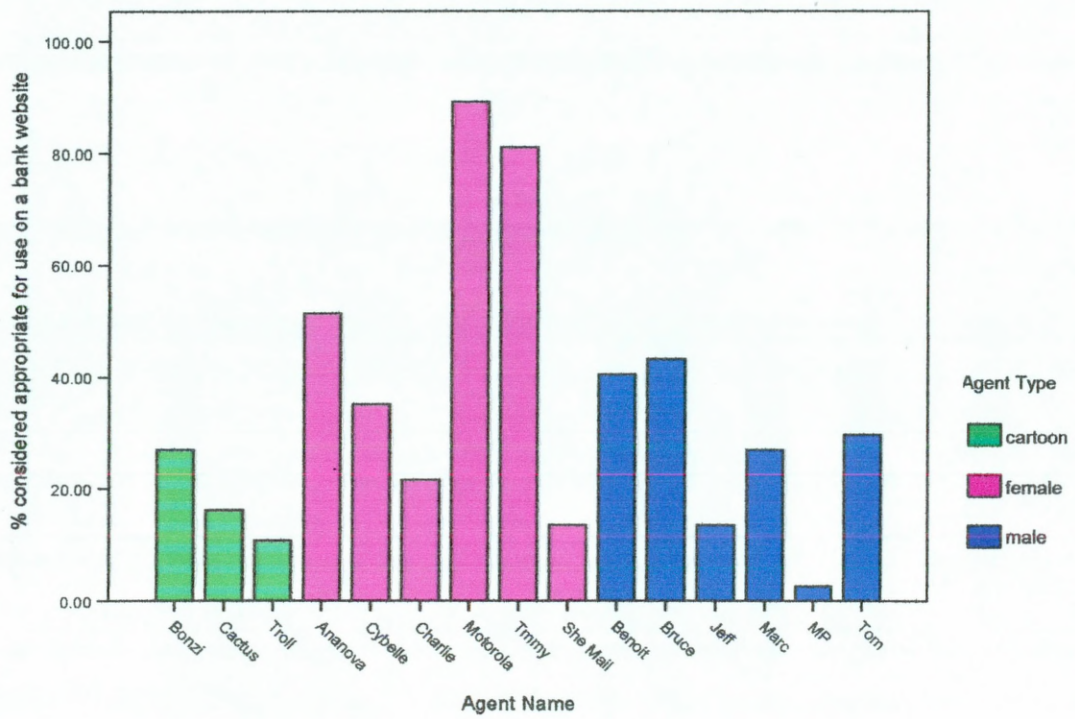


Figure 6.6: Agents considered appropriate for use on a bank website

6.4 Discussion

Initial analysis revealed that the influence of differing levels context (i.e. website levels) on each of the agent attractiveness categories was very small. Ratings of agents remained consistent, regardless of how 'good' or 'bad' the website it was shown in. It may have been the salience of the type of website used that was the driving factor influencing how participants rated the agents, rather than the actual design or 'look' of the website, as all website were bank/financial institutions. In order to investigate this further, the influence of context versus no context and imagined context was examined.

It was predicted that there would be an effect of context on rating of agent attributes, and Mann-Whitney analysis revealed that agents in the context condition were rated significantly lower on over half of the attributes measured, when compared to the imagined context and no context conditions. However, the difference between the imagined context and no context conditions revealed only differences on one of the attributes (memorability), with agents in the imagined context condition perceived as being less memorable. These finding suggest that embedding an agent in context has an adverse effect on judgements made. Thus, it is imperative that agents evaluated in future studies be shown in context in order to obtain a true reflection on how they will be perceived on the Internet.

In relation to the 'gender' of the agent, it was predicted that rating of agent attributes would be affected by agent 'gender'. Results showed that, regardless of context condition, there was a general positive regard for female agents. When compared to male agents and cartoon agents, female agents were perceived as being more attractive, trustworthy, sensible, useful and intelligent. Female agents were also considered to be more appealing and likeable when compared to male agents. The condition in which the agents were to be judged also had an influence on differences in attribute ratings between male, female and cartoon agents. Female agents were rated significantly lower on seven out of the nine attributes in the context condition when compared to the imagined context condition. Similarly, female agents were rated significantly lower on eight out of the nine attributes in the no context condition when compared to the imagined context condition. However, there were no

differences in attribute ratings for female agents between the context and no context condition. This suggests, for female agents, that there is a strong influence of implied context on user's perceptions, with female agents in the imagined context condition being perceived more favourably. Showing a female agent in context, or showing it independent of any context seems to have a detrimental effect on user's perceptions. However, it may be the case that an "imagined" context simply enhances perceptions of a female agent as participants may have been envisaging it being used in a website that they considered 'appropriate'.

With regard to male agents, context also had a detrimental effect of participants' perceptions. When compared to the no context condition, male agents in the context condition were rated significantly lower on five out of the nine attributes. Similarly, male agents were rated significantly lower on three out of the nine attributes in the context condition when compared to the imagined condition. There were no differences for male agents between the imagined condition and the no context condition. These results suggest, for male agents, that there is a strong influence of context on user's perceptions, with male agents in the context condition being perceived less favourably.

There were very few differences between the context, imagined context, and no context conditions for cartoon agents, suggesting that perceptions of cartoon agents remain stable, regardless of the influence of context.

Agent attractiveness

It was predicted that agent attractiveness would affect rating of agent attributes, and results showed that, regardless of context condition, there was a general positive regard for attractive agents. When compared to unattractive agents and agents of average attractiveness, attractive agents were perceived as being more attractive, appealing, trustworthy, friendly, sensible, useful, intelligent, and likeable. The only attribute on which attractive agents were judged lower was on memorability, when unattractive agents were judged to be significantly more memorable.

The condition in which the agents were to be judged also had an influence on differences in attribute ratings between attractive, unattractive and average agents. Attractive agents were rated significantly lower on seven out of the nine attributes in the context condition when compared to the imagined context condition. Similarly, attractive agents were rated significantly lower on five out of the nine attributes in the no context condition when compared to the imagined context condition. However, attractive agents were judged to be significantly less attractive in the context condition compared to the no context condition. This suggests, for attractive agents, that there is a strong influence of implied context on user's perceptions, with attractive agents in the imagined context condition being perceived more favourably. Showing an attractive agent in context, or showing it independent of any context seems to have a detrimental effect on user's perceptions.

Similar results were found with regard to agents of average attractiveness, with context having detrimental effect of participants' perceptions. When compared to the imagined context condition, average agents in the context condition were rated significantly lower on seven out of the nine attributes. Similarly, average agents were rated significantly lower on three out of the nine attributes in the context condition when compared to the no context condition, and significantly lower on two out of the nine attributes in the no context condition compared to the imagined context condition. These results suggest, for both attractive agents and agents of average attractiveness, that there is a strong influence of context on user's perceptions, with agents in both categories being perceived less favourably in the context condition.

There were very few differences between the context, imagined context, and no context conditions for unattractive agents, suggesting that perceptions of unattractive agents remain stable, regardless of the influence of context. These are similar to the finding for cartoon agents, and it may be the case that both cartoon agents and unattractive agents are perhaps considered inappropriate for use on a bank website, and therefore the ratings are low, regardless of the context in which the agents are shown.

The influence of context, both in terms of the 'gender' and attractiveness of the agent somewhat supports the findings of Eagly et al. (1991), who suggested that context

influenced judgements of other people. Context had a detrimental effect on judgements of female agents, male agents, attractive agents, and agents of average attractiveness when compared to the imagined context condition. Similarly, context had a detrimental effect on judgements of male agents, cartoon agents, attractive agents, unattractive agents, and agents of average attractiveness when compared to the no context condition. However, Eagly and colleagues suggest that the evaluation of the actual context itself may be an influential factor. This was not found in the current study, where the design of the website did not affect judgements of the agents. These findings do, however, support the findings of Larose, Tracy and McKelvie (1993), who found that the favourability of the magazines used in their study did not influence judgements made about the target persons.

Appropriateness analysis

Participants in the context and imagined context conditions were asked whether they thought each agent was 'appropriate' for use on a bank website. Analysis showed that there was no significant difference between context and imagined context when deciding if an agent was appropriate. Overall, female agents were considered to be most appropriate, followed by male agents. Cartoon agents were considered to be least appropriate for use on a bank website. In order to investigate if there were any differences in attributes between the two agents considered most appropriate for use on a bank website, Mann Whitney analysis was carried out. It was found that there were significant differences between Motorola and Tmmy on six out of the nine attributes. Tmmy was rated as being significantly more attractive, memorable, and appealing when compared to Motorola. However, Motorola was rated as being significantly more sensible, useful, and intelligent when compared to Tmmy. These findings suggest that not only are 'usability' factors such as usefulness and intelligence important in determining if an agent is appropriate, but aesthetic factors such as attractiveness and appeal are perhaps equally important. However, both of these factors could contribute to the overall view of attractiveness. As Solomon, Ashmore and Longo (1992) point out, there are different types of attractiveness, and it may be the combination of 'usability' and 'aesthetic' factors that determined participant's views on appropriateness. Similarly, Langmeyer and Shank (1994) found that products were considered "beautiful" based on they way they worked,

functionality, and how well deigned they were. They also found that personality and intelligence were important aspects of beauty. Their main conclusion was that although physical appearance is the most salient aspect of initial impressions of a person or product, beauty is more than skin deep. The findings from the current study suggest that these finding may also extend to agents, particularly when considering their appropriateness to a given context. Although there were significant differences between Motorola and Tmmy on six out of the nine attributes both of these agents were rated consistently high on all nine attributes. On four out of the nine attributes, Motorola was rated highest (trustworthiness, sense, usefulness, and intelligence), and was rated in the top four on attractiveness, appeal, and likeability. Tmmy was rated in the top 3 on seven out of the nine attributes (attractiveness, appeal, trustworthiness, sense, usefulness, intelligence and likeability).

However, from the current study, it is unclear whether these perceptions of appropriateness were based on the agents shown, or on participants views that a female agent would, stereotypically, be best suited to that role. Thus, the next series of studies will examine occupational stereotypes and, subsequently, if these stereotypes extend to virtual agents.

Chapter 7 : The role of occupational stereotypes: evidence for gender biases?

7.0 Aims of chapter

The previous experimental chapters suggested that a number of factors may critically influence judgements of agent suitability and a number of hypotheses were considered. The intrinsic characteristics of agents such as perceived attractiveness and gender, and inferences regarding personality and intelligence could significantly influence expressed preferences for agents. However, the preference may also reflect gender or role specific inferences about suitable qualities, based on real-world experience, in the form suggested by Reeves and Nass (1996). In addition, the findings in Chapter 6 suggest that context has a significant impact on perceptions of attractive agents and agents of average attractiveness. However, the occupation suggested by the context (a financial services website) made it difficult to determine whether female agents were perceived to be most appropriate for the occupation due to the general positive regard for female agents, or whether the occupational stereotypes were driving these judgements.

The aim of this Chapter is to determine the extent to which stereotypes exist about a set of occupations and to identify whether these stereotypes include a gender bias. Subsequent Chapters will then examine whether these stereotypes extend to virtual agents.

7.1 Introduction

There are many occupations that are almost totally dominated by one gender (for example, mechanics and dental hygienists). Eagly (1987) theorised that gender stereotypes are shaped by observing men and women in their roles in everyday life. She argued that men and women are perceived as separate social groups, and both groups are deemed to possess attributes appropriate to the roles they occupy. Similarly, Glick (1991) found that female dominated occupations were perceived to require a higher level of feminine personality traits for job success. In a later study,

Cejka and Eagly (1999) investigated the importance of personality traits, along with physical attributes, and cognitive characteristics on perceived occupational success. They predicted that gender stereotypes would lead to masculine characteristics being associated with higher prestige, higher income occupations. Their results showed that, in general, the most important dimension was masculine cognitive ability, followed by feminine cognitive ability, then feminine and masculine personality characteristics. Feminine physical characteristics were least important in occupational success. The ordering of these dimensions changed when considering gender-specific occupations. For male dominated occupations, masculine cognitive characteristics were most important, followed by feminine cognitive, masculine personality, masculine physical, feminine personality, and feminine physical characteristics. For female dominated occupations, the most important attributes were feminine personality, followed by masculine cognitive, feminine cognitive, masculine personality, feminine physical, and masculine physical characteristics.

Occupations were thought to be more prestigious if they required masculine personality or cognitive characteristics. Masculine personality characteristics such as competitiveness, dominance, and aggressiveness were also related to perceived earnings. That is, occupations that were well paid were those believed to require masculine qualities of personality, and male dominated occupations were associated with higher wages. In fact, the strongest predictor in assessing gender division in occupations was feminine qualities of personality. Success in female-dominated occupations was associated with qualities such as gentleness, helpfulness, sociability, kindness, and supportiveness. Similarly, success in male-dominated occupations was associated with masculine qualities of personality.

Glick, Wilk and Perreault (1995) suggested that occupational stereotypes are based on images of people holding a particular job, and include factors such as gender, lifestyle and personality traits, rather than the job itself. Although it has been demonstrated that people have relatively complex ideas about what different jobs involve, Gottfredson (1981) (cited in Glick et al., 1995), argues that people automatically classify occupations according to two dimensions: gender-type and prestige. These dimensions not only determine how people classify jobs, but also determine a person's vocational choice, with few people choosing to pursue a career in which the

gender-type or prestige does not match with their own gender or socio-economic background. Therefore, this may be one reason for the continued segregation and stereotypes that still exist in the job market. Although "gender-type" is usually treated as a uni-dimensional construct, other researchers (e.g. Spence, 1993) have suggested that it can be multi-dimensional, and can include a number of factors, such as personality and physical traits. Although these factors have been found to be related to gender segregation in occupations, most research has found that between 50% and 70% of the variance in subjective masculinity-femininity ratings in occupations was due to sex ratios (i.e the number of males and females in a particular job), and that these sex ratios are almost identical to the perceived 'gender-type' of an occupation. This again suggests that people's views on occupations are based upon who holds a particular job, rather than the skills or traits required to do the work.
However, Shinar (1975) (cited in Glick et al., 1995), argued that there are three potential factors of masculinity-femininity ratings: sex ratio, personality traits (masculine and feminine), and physical characteristics (masculine and feminine), and that ratings on each of these three factors were highly correlated. Glick (1991) further examined these factors and found that there were only moderate correlations between sex ratios and masculine and feminine personality traits. He also found that while masculine and feminine traits were positively correlated with occupational prestige ratings, sex ratios did not predict prestige ratings. Glick suggested that these finding supports the idea of treating 'gender-type' as a multi-dimensional construct, and that these three factors should be taken into consideration rather than a single "masculine-feminine" rating. In their study into gender and status occupational stereotypes, Glick et al. (1995) corroborated the findings of Gottfredson (1981) and found that people classified occupations according to two main dimensions: prestige/intelligence and gender-type. They stress that gender-type is not synonymous with sex ratio, but also represents the physical 'type' associated with an occupation, with "pink collar" occupations requiring feminine personality traits (such as nurse), at one end of the pole, and "blue collar" occupations requiring masculine physical traits (but not masculine personality traits) at the other end of the pole.

Although Glick et al. did find a relationship between sex ratio and prestige, the relationship was not linear; there were a high number of males in both the high and low status jobs. Another interesting factor to emerge from this study was the

importance of masculine personality traits on communication skills. However, further examination of the occupations revealed that occupations such as lawyer and 'armed forces' scored highly on communication, suggesting that communication skills (in this case) relate to attributes such as persuasion and leadership.

Occupational stereotypes may have implications when choosing an agent as the context (or occupation) in which the agent is placed may determine which characteristics users will attribute to it. In addition, there may be characteristics which are more, or less, salient, depending on the context, therefore the influence of these different attributes must be determined. In other words, the effects of occupational stereotypes could have a major influence on choosing suitable agents for use on the Internet. If an agent chosen is inconsistent with a user's stereotypical views of the occupation it is portraying then the agent may more easily be associated with incompetence and less helpful attributes than if it were consistent. If inconsistencies and stereotype violations occur, people are harsher critics and a gender inappropriate agent may need to be excellent to overcome this. On the other hand, agents that are consistent with the context (or occupation) may be seen more favourably and judged to be more competent. Cann (1993) suggested that competence is linked to gender-consistency in occupations, therefore agents that are considered appropriate for an occupation may, in turn, be considered more competent.

However, there is a possibility that occupations are now no longer so strongly perceived as male or female gender biased. In addition, in a climate of political correctness, participants may be unwilling to endorse gender biases in questionnaire responses. The following studies were, therefore, designed to investigate whether overt judgements about agent efficacy might be affected by gender-bias with respect to the occupational roles portrayed.

7.2 Study 1

The experimental studies carried out so far have shown that embedding an agent in context may have a detrimental effect on judgements made, specifically for female agents. Thus, it is imperative that an agent chosen for use on a service provision website be appropriate for the context. Results from earlier studies showed a general

positive regard for female agents where a financial service provision role was identified for the agent. There are a number of reasons why this may have been the case, but the main reason may simply be expectations, that is, occupational stereotypes. People may simply expect an agent on a financial service provision website to be female, since the majority of people employed as bank tellers are female.

The current study aimed to identify people's stereotypical views about a number of different occupations and to determine whether these stereotypes included a gender bias. It was hoped that by determining which occupations were considered to be dominated by one gender that this information could be used to test the matching hypothesis, whereby those agents selected as appropriate for use on a website would either match the occupational role or not.

A within subjects design was employed in this study. The independent variable was the gender of the occupation, which had 3 levels – male, female, and neutral. The dependent variables were the responses to each of the 13 questions.

7.2.1 Method

7.2.1.1 Participants

A total of 59 participants took part in this study; 11 were male and 48 were female. They were recruited from Abertay University by means of convenience sampling.

7.2.1.2 Materials

15 occupations were chosen for this study – 5 'male' occupations (e.g. mechanic), 5 'female' occupations (e.g. florist), and 5 'neutral' occupations (e.g. estate agent). These were chosen on the basis of the number of males and females in each occupation according to the 1991 census data (see Appendix 27 for list of occupations).

A questionnaire was administered for each occupation, and contained 13 questions on factors such as the importance of trust, intelligence and attractiveness for each occupation (see Appendix 28 for questionnaire).

7.2.1.3 Procedure

Participants were asked to complete a questionnaire for each of the 15 occupations. The order in which the occupations appeared in the questionnaires was randomised. The task was to answer 13 questions relating to the particular occupation. The questions related to the gender division in the occupation, how well paid the occupation was, the age range of people in the occupation, how important trustworthiness, intelligence, quality of advice, supportiveness, and attractiveness were for the occupation, and how prestigious the occupation was. The final three questions related to the participant's personal preference on how important attractiveness, intelligence and supportiveness were for each occupation.

7.2.2 Results

For the purposes of this study, an occupation was deemed to have been judged 'male' if participants believed that 60% or more of people engaged in it were male and similarly for 'female' occupations.

Analysis of the gender division data revealed that eight out of fifteen occupations were rated as being 'female' (see Figure 7.1). Of these eight occupations, five were the original 'female' occupations (according to the census data used) - florist, bank teller, hairdresser, nurse, and shop assistant, and the remaining three were originally classed as neutral occupations (weather forecaster, counsellor, and holiday rep). Four out of fifteen occupations were considered 'male'. These were mechanic, lawyer, bank manager, and financial advisor. The remaining three occupations were considered 'neutral' (fitness instructor, estate agent, and doctor (which was originally classed as a 'male' occupation)).

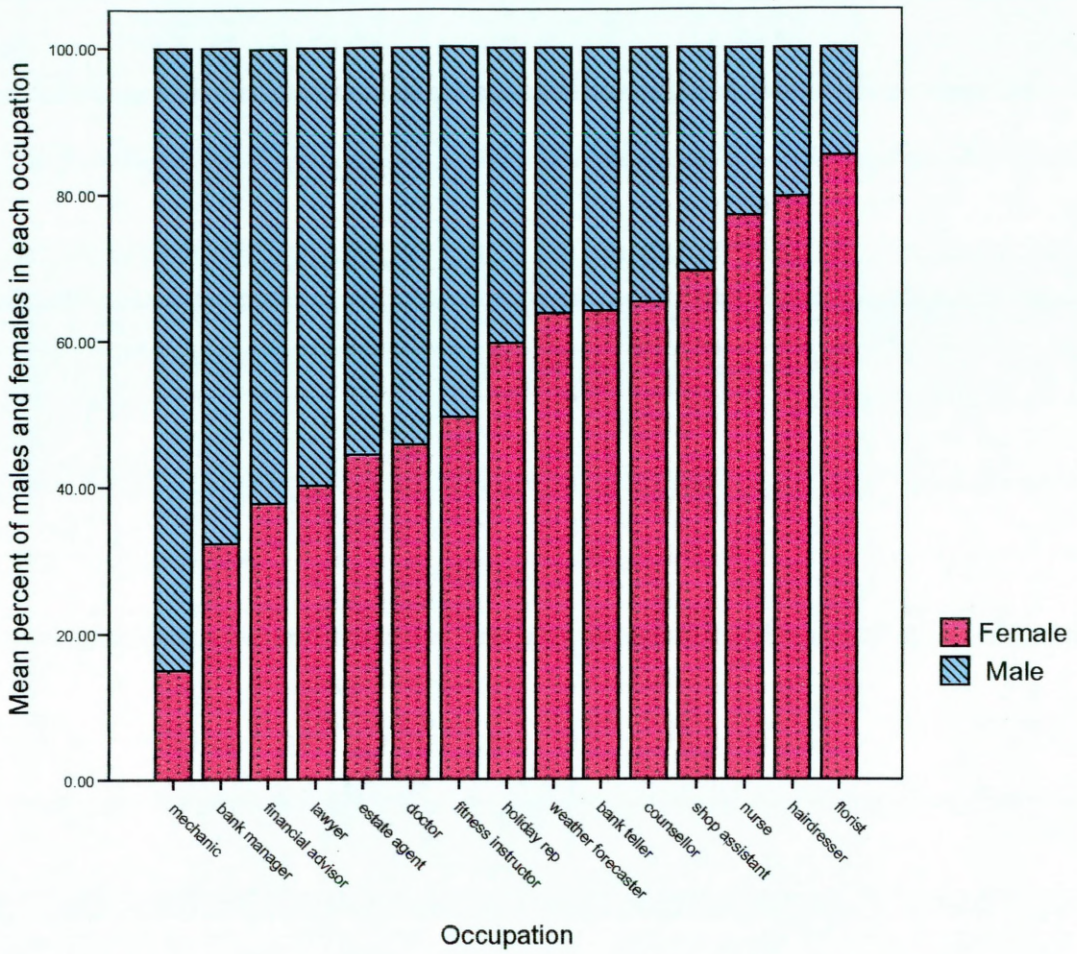


Figure 7.1: Perceived Gender distribution in each occupation

Differences between male and female occupations were analysed using a Mann Whitney test and revealed that there were significant differences on all attributes, with the exception of support and preferred support. Male occupations were rated higher on all attributes when compared to female occupations, with the exception of attractiveness and preferred attractiveness (see Appendix 29 for Mann Whitney statistic summaries).

The attributes of advice, attractiveness, intelligence, prestige, support, well-paid, preferred attractiveness, preferred intelligence, and preferred support were analysed using Multidimensional Scaling (MDS). This allowed each occupation to be compared against every other occupation on each attribute.

The MDS plots for each attribute are shown in Appendix 30. The MDS solutions were interpreted by visual inspection. In each case (with the exception of attractiveness and preferred attractiveness) one dimension was optimal for explaining the variance between occupations. On one end of the dimension were 'manual' occupations (such as mechanic and shop assistant) and on the other end 'professional' occupations (such as lawyer and doctor).

A multiple regression was performed to determine the best predictors of how prestigious an occupation was. The model predicting Prestige included nine predictors (advice, attractiveness, intelligence, support, trust, well-paid, preferred attractiveness, preferred intelligence, and preferred support), and accounted for a significant amount of variance [adjusted R-square = 0.692; $F(9,884) = 221.51$, $p<0.0001$]. The predictors Well-paid ($t=19.37$, $p<0.0001$), Intelligence ($t=5.72$, $p<0.0001$), Support ($t=2.58$, $p<0.05$), and Preferred Support ($t=2.32$, $p<0.05$) all had a significant impact on predicting how prestigious an occupation was. The impact of the predictors Trust, Advice, Attractiveness, Preferred Attractiveness, and Preferred Intelligence did not reach significance. The most important predictors were Well-paid (ß=0.483), followed by Intelligence (ß=0.231).

7.2.3 Discussion

The multidimensional scaling analysis between occupations on each of the attributes revealed that there was one main dimension on all attributes (with the exception of attractiveness and preferred attractiveness). Participants discriminated between occupations according to whether they were manual or professional. With respect to attractiveness and preferred attractiveness, occupations were grouped according to how 'social' the occupation was, with occupations such as holiday rep, weather forecaster, hairdresser (all female occupations), and fitness instructor (neutral occupation), accounting for one end of the dimension and the other occupations at the other.

A Mann-Whitney analysis revealed that there were significant differences on ratings between male and female participants only for the attributes of advice, attractiveness, intelligence, preferred attractiveness and preferred intelligence.

Female participants preferred mechanics to be more physically attractive than male participants did, whereas male participants preferred nurses and hairdressers to be more physically attractive than female participants did. Male participants also rated the importance of the physical attractiveness of hairdressers significantly more highly than female participants. Female participants rated the importance of the quality of advice from hairdressers significantly more highly than male participants. They also rated the intelligence of financial advisors more highly than male participants did, and preferred hairdressers and financial advisors to be more intelligent. These differences highlight that the stereotypes participants have regarding certain occupations are gender-specific.

However, as these were only small differences, male and female judgements were analysed together and the overall mean scores on the dimensions considered for the multiple regression analysis. Occupational Prestige was taken as the dependent variable, and the nine other category scales (advice, attractiveness, intelligence, support trust, well-paid, preferred attractiveness, preferred intelligence, and preferred support.) as predictor variables. Analysis showed that the best predictor for occupational prestige was how well paid the occupation was, followed by the importance of intelligence for the occupation. The regression analysis indicated that well-paid, intelligence, support, and preferred support all had a significant impact on predicting how prestigious an occupation was.

The relationship between prestige and how well paid an occupation is supports the findings by Cejka and Eagly (1999) and indicated that male dominated occupations are better paid and thought to be more prestigious compared to female dominated occupations. Mann Whitney analysis revealed that there were significant differences between perceived male and female occupations on all attributes measured, with male occupations rated higher on all attributes with the exception of attractiveness and preferred attractiveness.

The main limitation with this study was that the choice of occupations might have been too restricted. It may, for instance, be the case that the male occupations were considered more prestigious than female occupations simply because there were no prestigious female occupations included in the sample and it was certainly the case that the number of occupations considered by the participants to be 'male' was very low. In addition, the participant sample was predominately female. Although researchers such as Gottfredson has found no differences in gender when considering occupational stereotypes, that cannot be assumed in the current study as small but significant differences were found between male and female participants. Therefore, a second study was designed to examine stereotypes further but also to expand the number and type of occupation tested and to try to achieve a better gender balance across participants.

7.3 Study 2 Pilot phase

7.3.1 Method

7.3.1.1 Participants

10 participants took part in this study – 5 males and 5 females. They were all recruited from Abertay University by means of convenience sampling.

7.3.1.2 Materials

98 occupations were randomly chosen from a number of different sources. These included occupations used by Glick, Wilk and Perreault (1995) in their study of images of occupations, US Department of Labor, and the Occupational Outlook handbook. The occupations chosen for this study were selected to represent a wide range of prestige and gender-types (see Appendix 31 for full list of occupations). The occupations were listed on A4 paper, with a corresponding 7-point likert scale, on which 1 represented male dominance, and 7 represented female dominance.

7.3.1.3 Procedure

Participants were asked to rate for gender domination each of the 98 occupations on the scale. The order in which the occupations appeared was randomised for each participant. Each occupation was listed, with a 1 – 7 bi-polar likert scale to the right of it. Participants were asked to mark on the scale how much they thought an occupation was gender dominated, where 1 indicated that they thought the occupation was completely male dominated and 7 indicated that they thought the occupation was completely female dominated.

7.3.2 Results and Discussion

In order to test for differences on ratings of gender domination between male and female participants, Mann Whitney analyses were carried out. Results showed that, for each individual occupation, there were differences between male and female participants on only 4 out of the 98 occupations (child care, cashier, private investigator, and gynaecologist). In addition, across all occupations taken as a whole, there were no significant differences on ratings by male and female participants ($U = 119588.5$, $n1 = 490$, $n2 = 490$, $p = 0.915$), therefore ratings were analysed collectively.

Out of the 98 occupations sampled, 19 were considered to be strongly female dominated and 12 were strongly male dominated. An occupation was considered strongly gender-dominated if the mean occupation score was more than 1 standard deviation above (for female) or below (for male) the overall mean. It was decided to change the criterion for deciding if an occupation was gender-dominated as there were significantly more occupations in this study, and therefore it was felt that the criterion should be more stringent. The female-dominated occupations were receptionist, nursery teacher, librarian, secretary, childcare worker, cashier, dental hygienist, housekeeper, bank teller, nurse, dietician, legal secretary, florist, cabin crew (flight attendant), shop assistant, hairdresser, travel agent, midwife, and beautician. The male dominated occupations were chemical engineer, flight engineer, electrician,

painter/decorator, butcher, labourer, carpenter, miner, farmer, funeral director, car dealer, and builder.

7.4 Study 2 – Occupational Stereotypes

The jobs generated in the pilot phase were used to examine occupational stereotypes by means of a job images questionnaire. The questionnaire was based on the job images questionnaire developed by Glick et al. (1995) and aimed to investigate the structure of occupational images.

7.4.1 Method

7.4.1.1 Participants

64 participants took part in this study – 28 were male and 36 were female. The participants were recruited from Abertay University and Dundee College by means of convenience sampling.

7.4.1.2 Materials

Eighteen occupations (see Appendix 32) were selected from the pilot study, six male occupations (e.g. labourer, car dealer), six female occupations (e.g. nursery teacher, cabin crew) and six neutral occupations (e.g. police officer, lawyer). A questionnaire was used (see Appendix 33) for each of the 18 occupations, each with 30 questions: 10 questions relating to attributes of the job/jobholder (extent to which job is male or female dominated; average age of typical jobholder; extent to which typical jobholder possesses masculine and feminine physical traits; extent to which physical strength is required for the job; degree of intelligence of the typical jobholder; educational attainment of typical jobholder; prestige of the job; trustworthiness of typical jobholder; attractiveness of typical jobholder), 8 questions relating to specific skills that may/may not be required in a particular occupation (these included verbal, analytical, mechanical, spatial, clerical, and communication abilities), and 12 questions relating to the degree to which the job requires masculine, feminine, and

neutral personality traits (selected from Bem Sex Role Inventory (Bem, 1974)). Each question was rated on a 7-point likert scale.

7.4.1.3 Procedure

Each participant was given a set of questionnaire booklets and asked to answer each question for each of the 18 occupations. The order in which the occupations appeared was randomised.

7.4.2 Results

For the primary analysis, participants' ratings were averaged, as the appropriate unit of analysis was by occupation, not participant. In order to verify that participants' ratings could be legitimately averaged, two types of analyses were conducted. Firstly, reliability analysis was conducted for each attribute rated, and secondly, possible gender differences between participants were examined.

Reliability Analysis

Separate reliability coefficients were calculated for each of the attributes rated. To compute the coefficients, a separate data file was created for each attribute rated (e.g. degree of intelligence required). Each line of the data file included ratings of one specific occupation by all participants (i.e. occupations were treated as participants usually are, with each line of data containing ratings of a specific occupation on a single attribute). The data entered on each line were the ratings of the occupations on one specific attribute by each participant. Analysis revealed that the ratings were highly reliable. Alphas ranged from .89 to .99 (average alpha = .97). The high reliability suggests that the different raters had highly consistent views about whether particular occupations were high or low on each of the attributes rated.

Gender differences

In order to test for differences between male and female participants, a Mann-Whitney analysis was carried out. Results showed that there were significant differences between male and female participants on 22 of the 30 attributes measured. However, correlational analysis of the average ratings by male and female participants revealed an extremely high significant correlation (r=.97), which is represented graphically in Figure 7.2 below.

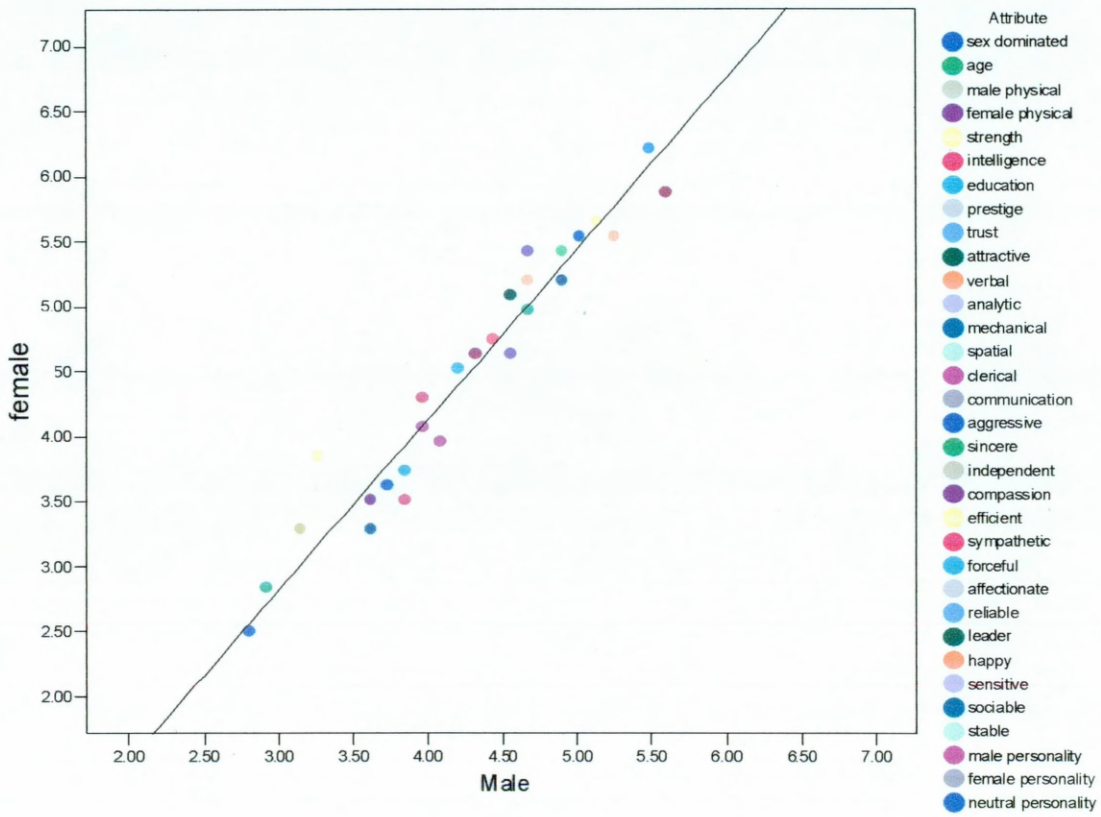


Figure 7.2: Correlation of male and female participant's mean rating on each attribute

Because of the high degree of agreement among participants, the ratings on each occupation were averaged across raters to give a single rating on each attribute for each occupation.

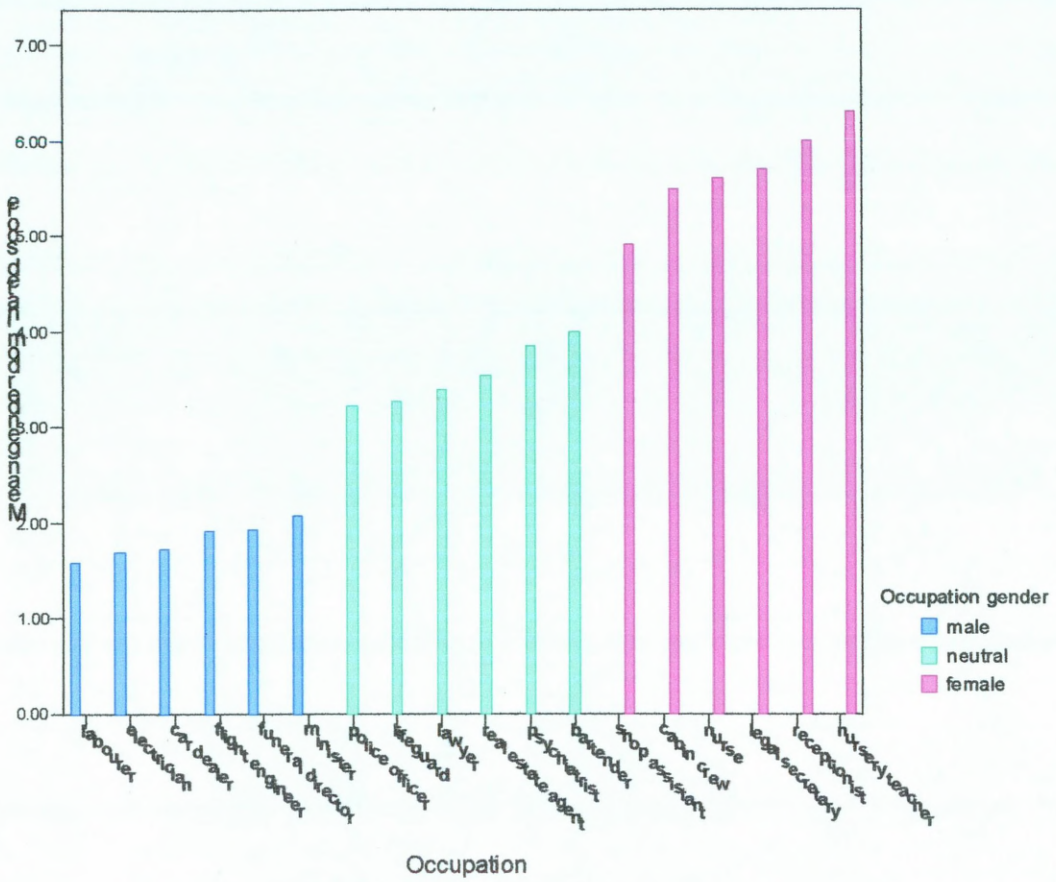


Figure 7.3: Gender domination of each occupation

The mean gender domination score (see Figure 7.3) for each occupation revealed that the occupations were rated as consistently male, female, or neutral, as per the pilot study results.

Factor Analysis

Factor analysis was used to determine the underlying dimensions on which the ratings of occupations were clustered. Principal components factor analysis with a varimax rotation yielded a five-factor solution (using eigenvalue of 1 as the cut-off) that accounted for 93% of the variance. Table 7.1 shows the factor loadings for the variables. The first two factors accounted for the majority (67%) of the variance. Factor 1 was interpreted as emotionality dimension, with sympathetic, compassionate, affectionate, sensitive, stability, sincerity, trustworthiness, and reliability showing the highest factor loadings. Factor 2 was interpreted as a gender dimension, with happiness, communication skills, sociability, verbal and clerical skills, showing high

factor loadings, and strength, spatial skills, male physical traits, and mechanical skills showing high negative loadings. Factor 3 represented a prestige/intelligence dimension, with high factor loadings on the attributes of prestige, independence, education, intelligence, efficiency, and analytical skills. Factor 4 represented an attractiveness dimension, with female physical traits, sex dominated (female dominated), and attractiveness having high factor loadings, and age showing a high negative loading (i.e. as age increases, attractiveness decreases). Factor 5 represented a forcefulness dimension, with leadership abilities, forcefulness, and aggressiveness showing high factor loadings.

Table 7.1

Rotated Component Matrix for 30 Attributes

Variable	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Sympathetic	**0.94**	0.27	0.15	0.00	-0.06
Compassionate	**0.92**	0.28	0.23	-0.02	-0.01
Affectionate	**0.92**	0.27	0.06	0.06	-0.14
Sensitive	**0.90**	0.34	0.10	-0.02	-0.05
Stable	**0.84**	0.20	0.47	0.01	0.11
Sincere	**0.83**	0.33	0.37	-0.19	-0.09
Trustworthiness	**0.80**	-0.05	0.50	0.19	-0.13
Reliable	**0.65**	-0.37	0.57	0.11	-0.22
Communication	0.38	**0.84**	0.28	0.06	0.21
Verbal	0.32	**0.84**	0.36	0.00	0.19
Sociable	0.37	**0.70**	-0.35	0.34	0.21
Clerical	-0.09	**0.66**	0.56	0.04	-0.27
Happy	0.49	**0.62**	-0.30	0.39	0.01
Mechanical	-0.45	**-0.73**	0.08	-0.22	0.10
Male physical	-0.28	**-0.78**	-0.08	-0.28	0.43
Spatial	-0.16	**-0.80**	0.34	-0.07	0.23
Strength	0.02	**-0.90**	-0.26	0.13	0.20
Analytic	0.07	0.04	**0.92**	-0.09	0.26
Intelligence	0.28	0.11	**0.89**	-0.16	0.22
Education	0.31	0.11	**0.88**	-0.08	0.22
Prestige	0.42	0.04	**0.81**	-0.10	0.36
Efficient	0.25	0.06	**0.79**	0.39	-0.26
Independent	0.32	0.00	**0.70**	-0.24	0.49
Attractiveness	0.00	0.16	0.03	**0.95**	0.08
Sex dominated	0.18	0.52	0.06	**0.76**	-0.28
Female physical	0.24	0.62	0.00	**0.66**	-0.32
Age	0.25	0.14	0.27	**-0.90**	-0.06
Leadership	0.51	-0.03	0.52	-0.16	**0.61**
Aggressive	-0.28	-0.18	0.19	0.00	**0.85**
Forceful	-0.14	-0.12	0.29	0.02	**0.91**

A two-dimensional map of the occupations plotted on the emotionality and gender factors is presented in Fig. 7.4. It can be shown from this plot that the gender factor does not simply represent how male or female dominated an occupation is, rather it shows the traits associated with an occupation. The 'masculine' end of the scale represents occupations dominated by males, which are also associated with male physical traits, physical strength, mechanical and spatial skills. On the 'feminine' end of the scale, occupations are associated with feminine personality traits, and include occupations where a high degree of communication and verbal skills are important.

The masculine/low emotion part of the plot was male dominated, with three out of the six male occupations, and only one female occupation falling into this category. The feminine/high emotion part of the plot was mainly dominated by female and neutral occupations, with six out of twelve possible occupations falling into this category. Only one male occupation fell into this category (minister). There were no female occupations and only one male and one neutral occupation in the masculine/high emotion part of the plot. The feminine/low emotion part of the plot included two female occupations, two neutral occupations, and one male occupation. It is interesting to note that the occupations that fall in to the feminine/high emotion category are the occupations that require a high degree of interpersonal skills, such as nurse, police officer, and minister, whereas the masculine/low emotion category is dominated by 'manual' occupations such as labourer and electrician. This suggests that rather than a manual-professional bi-polar scale, it may be more beneficial to rate occupations on an interpersonal-manual scale.

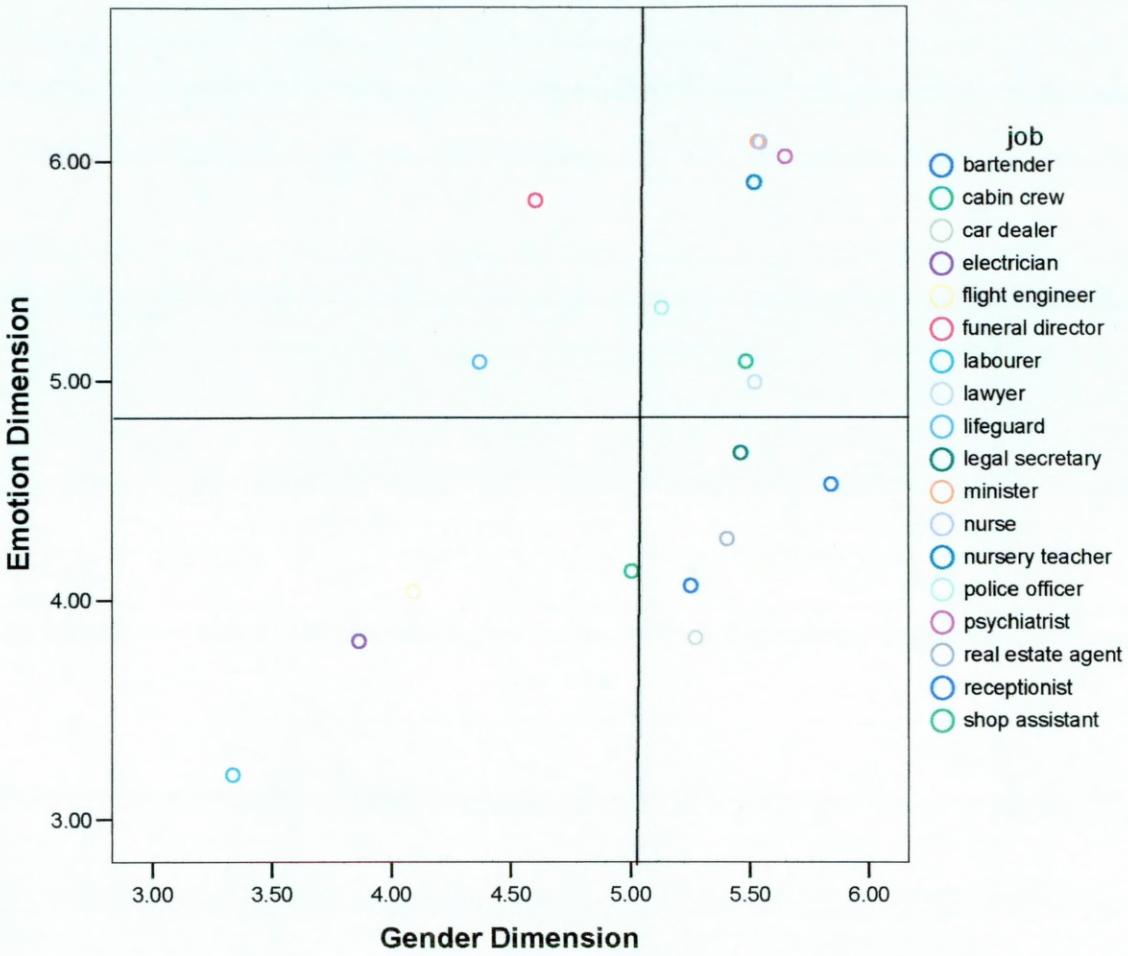


Figure 7.4: Map of occupations by Emotionality and Gender dimensions

7.4.3 Discussion

The aim of this study was to determine if occupational stereotypes exist, and, if so, the extent of these stereotypes. The main findings of this study suggest that occupational stereotypes do exist, but that the perceptions of occupations is more complex than simply considering the perception of the number of males and females in each occupation.

Although there were two main underlying dimensions found in this study, they were not the same dimensions as those found by Gottfredson (1981), and Glick et al.

(1995). They found that people automatically classify occupations according gender-type and prestige. While prestige was a factor in the findings of the current study, "emotionality" was found to be the most important dimension (which included attributes such as sympathy, compassion, affection and trust), along with gender-type, along which people discriminated between occupations.

However, the results of this study do support the findings of Glick et al., who also found that gender-type is not the same as sex ratio. The results showed that the masculine end of the gender-type dimension included attributes such as mechanical and spatial skills, as well as strength and masculine physical traits, whereas the feminine end of the scale included attributes such as communication and verbal skills, as well as 'sociable' and 'happy'.

When the eighteen occupations were plotted along the 'emotionality' and 'gender-type' dimensions, it can be seen that the gender factor does not simply represent how male or female dominated an occupation is, rather it shows the traits associated with an occupation. Twelve out of the eighteen occupations lie on the 'feminine' end of the 'gender-type' dimension, and although this includes all but one of the 'female' occupations (the exception being 'shop assistant' which only just falls into the 'masculine' end of the scale), it also includes five out of the six neutral occupations, and two of the 'male' occupations. Furthermore, seven out of the twelve 'feminine' occupations are also rated highly on the 'emotion' dimension, and are associated with feminine personality traits, and include occupations where a high degree of compassion and empathy is important, with occupations such as nurse, minister, and cabin crew falling into the feminine/high emotion category. Five out of the twelve 'feminine' occupations fall into the 'low' emotion category, but are also associated with feminine personality traits, and include occupations where a high degree of communication and verbal skills are important, with occupations such as receptionist, bartender, and car dealer falling into the feminine/low emotion category.

The 'masculine' end of the scale only includes six out of the eighteen occupations. These occupations are predominantly male dominated, but also include occupations which are associated with male physical traits, physical strength, mechanical and spatial skills. The masculine/high emotion category includes two occupation; funeral

director and lifeguard, both of which require a high degree of trust and reliability. Four out of the six occupations fall into the masculine/low emotion category, and are mainly 'manual' occupations, such as electrician and labourer.

The spread of occupations across the two dimensions may influence decisions made about the most appropriate agent to be used in that particular context. For example, it may not be simply a case of choosing a 'gender-consistent' agent for use in a particular context (or occupation), as inferences may be made about the attributes an agent has to possess in order for it to be 'suitable' for that role.

In addition, there may be characteristics which are more, or less, salient, depending on the occupation, and the influence of these different attributes must be determined. Users may want an agent to be consistent with their stereotypical views of someone in that occupation in order for them to be convinced of its competence. If the agent is considered inconsistent with the occupation and stereotype violations arise, users may view the agent more harshly and a gender/emotionally inappropriate agent may need to be exceptional to overcome this.

In short, matching of perceived gender specific role assignments or gender specific characteristics to the functional requirements of the occupation can and will occur in real life. The question remains, however as to whether such matching would be expected to occur in the case of virtual agents. However, even if it is likely to occur, if the aesthetic appeal of the agent is high, there are two possibilities: the aesthetic appeal dominates the judgements made and negates the effect of gender-bias in all cases or, alternatively this would only be the case with female agents as they have much higher attractiveness ratings. To test this hypothesis, roles strongly identified with one gender or another, and perceived agent suitability, would provide suitable cases for investigation.

Chapter 8 : Agent and occupation interaction

8.0 Aims of chapter

Although most literature on occupation selection suggests support for the 'what is beautiful is good' stereotype, some studies have found results that are contradictory to this. There may be other characteristics of an endorser that are more significant than physical attractiveness (Till and Busler, 1998). Gender, expertise, celebrity status, likeability, and similarity with target audience are all factors, besides physical attractiveness, that may have an effect on source credibility (Belch and Belch, 1998).

This chapter aimed to determine whether perceived agent qualities match-up with perceived occupation attributes, and to examine the influence of agent and occupation compatibility. It may be the case that there need not be agent/occupation compatibility and the 'beauty is good' hypothesis will provide a better explanation in that beautiful agents will always be the most favourably rated for all occupations.

8.1. Introduction

In a study investigating gender and occupations, Heilman and Saruwatari (1979) found that if a position to be filled was a traditionally male role, unattractive females were rated more favourably than attractive females. However, in sex-congruent jobs, attractiveness was considered an advantage, as attractive applicants were rated as being more qualified for the job than their unattractive counterparts. In a similar study, Shahani and Plumitallo (1993) found that in performance appraisal situations attractive employees were perceived as failing due to a lack of effort, whereas unattractive employees were though to fail due to bad luck. These results were consistent for male and female employees.

It has been suggested that in order for advertisements to be credible, and for consumers to relate to an attractive source, there has to be an association, or "match", between the endorser and the product (McCracken, 1989; Till and Busler, 1998). This

suggests that the 'best' choice of agent for a website might be one where it matched the perceived stereotype of the occupation it is portraying.

In Chapter 6, it was reported that, compared to male and cartoon agents, female agents were considered to be more appropriate for a financial service provision website. However, it did not clarify which, if any, of the agents users would most like to see and interact with on a website. It may be the case that the most helpful and trustworthy looking female agent would be judged the most appropriate for the role of a bank teller, but that the most attractive female agent (assuming that they are not the same) is the one that most people would prefer to see on the website. In other words, it is possible that aesthetics, rather than appropriateness, will be the driving factor in people's decisions of what they want in an agent, and attractive agents may be selected over and above others for websites. On the other hand, it may be that the expressed stereotypes for an occupation will predict the 'best' agents.

8.2 Current Study

This study aimed to match-up agents and occupations to determine what are the 'best' agents for a given set of occupations, and whether these agents varied across different occupations.

The use of physically attractive endorsers may engender considerable influence on consumer purchase intentions for low involvement products. Peripheral cues such as colours, and pictures of endorsers could attract more interest for purchases from low involved customers. Likewise, an attractive agent on a website may influence interaction and general positive regard if the task were low in self-relevance, for example information retrieval, but aesthetics may have little relevance if the task were high in self-relevance.

In order to assess whether there is a set of generic characteristics for agents that people would like in an agent, or would expect an agent to have, participants were asked to rate each individual agent on a number of attributes including attractiveness, trustworthiness and intelligence, to determine if certain aspects of their perceived make-up critically determined their selection for specific roles.

By asking participants to match each occupation with an agent, it was hoped that this would allow assessment of the interaction between agent and occupation characteristics, and to determine whether the most appropriate agents are more likely to be female (i.e. were female agents the best default option?). In addition, in order to assess if any differences exist between what is appropriate and what is preferred, participants were asked to match up which agent they would most prefer to interact with on a website and least like to interact with for each occupation.

Thus, there are two possible attributes that could determine the suitability of an agent to a website:
Attractiveness, if the 'what is beautiful is good' paradigm (Dion et al., 1972) is true, or Gender, if the stereotypes of the occupation are more salient gender may determine agent appropriateness if the occupation is stereotypically male or female. In addition to this, female agents were rated more favourably in previous experiments, thus it may be that participants will regard female agents as the best choice for an occupation as they are considered to have more positive attributes.

A within subject design was employed in this study. The independent variables were the gender of the agent, which had two levels – male and female; attractiveness of the agents, which had three levels - attractive, unattractive, and of average attractiveness; and type of occupation, which had four levels – high emotion/ female, low emotion/female, high emotion/male, and low emotion/male. The dependent variables were the ratings of each agent and occupation, the appropriateness of each agent to each occupation, and the agent preferred for each occupation.

It was hypothesised that the appropriateness of an agent to an occupation would vary as a function of agent gender. It was also hypothesised that attractive agents would be most preferred over and above all other agents, particularly for service, sales, or public relation related occupations.

8.2.1 Method

8.2.1.1 Participants

A total of 98 participants took part in this study, 50 of whom were male and 48 were female. Participants were recruited from the Abertay University and Dundee College by means of convenience sampling.

8.2.1.2 Materials

Occupations were selected on the basis of factor analysis from the occupational stereotypes study (Chapter 7). The factor analysis revealed 2 main dimensions – emotion (sensitive/insensitive) and gender (masculine/feminine). The factor 'gender' does not simply represent how male or female dominated an occupation is, rather it shows the traits associated with an occupation. The 'masculine' end of the scale represents occupations that are mainly dominated by males, but which are also associated with male physical traits, physical strength, mechanical and spatial skills. On the 'feminine' end of the scale, occupations are associated with feminine personality traits, but include occupations where a high degree of communication and verbal skills are important.

Eight occupations were selected on the basis of these dimensions – four male dominated, three female dominated, and one neutral occupation (lifeguard was rated as a neutral occupation but was included as there were no female occupations in the sensitive/male category). In order to avoid any bias in the prestige of occupations, only "manual" occupations (as classified in Study 3, Chapter 7) were selected. It should be noted that although "minister" was a male dominated occupation, it falls into the feminine/high emotion category. Similarly, "shop assistant" was a female dominated occupation and falls into the masculine/low emotion category, and "car dealer" was a male dominated occupation and falls into the feminine/low emotion category. The occupations chosen are shown in Table 8.1.

Table 8.1

Range of Eight Occupations

	Gender	
Emotionality	Masculine	Feminine
High	Funeral director	Minister
	Lifeguard	Cabin crew
Low	Shop assistant	Car dealer
	Labourer	Receptionist

A set of twelve pre-selected agents (from Experiment 6) was also used (see figure 8.1). These agents were selected as they were found to differ significantly in levels of attractiveness, and included three attractiveness groups: 'attractive'; 'of average attractiveness'; and 'unattractive'. Within each of these attractiveness groups there were two male agents and two female agents.

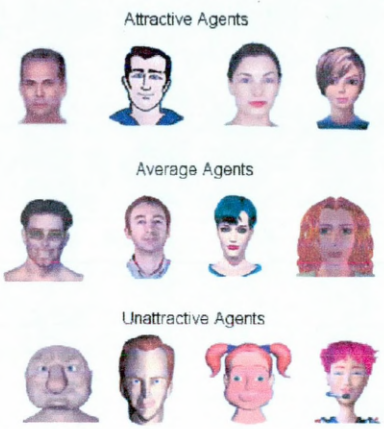


Figure 8.1: Attractive, average, and unattractive agents

Ratings task

An attributes questionnaire was used for rating each of the 12 agents, which consisted of an image of an agent at the top of the page, with seventeen attributes listed below on a 7-point bi-polar scale. Participants were instructed to rate the agent on each of the attributes, based purely on physical appearance (see Appendix 34). A similar questionnaire was used to rate the importance of each of the seventeen attributes for each occupation. Participants were instructed to rate the importance of each attribute on a 1-7 scale, where 1 was 'unimportant' and 7 was 'important' (see Appendix 35).

The attributes chosen were based on the job images questionnaire used in the previous study, and included attractiveness, trustworthiness, intelligence and strength.

Matching task

The instructions given to participants informed them that they were to make judgements about a number of different occupations and agents. In addition, a definition of an agent was provided ("An agent is a character that is used on a website with which users can interact"). The matching task required participants to think about each occupation and then match each one with an agent, for a given a criterion. The criteria were: *most appropriate, least appropriate, prefer to interact with, and least like to interact with.* For example: "Please match each occupation with the agent you think would be most appropriate in that role. You may select an agent for more than one occupation" (see Appendix 36).

8.2.1.3 Procedure

Ratings task

Participants were given a set of 12 questionnaires, one for each agent, and asked to rate the agent shown on each of the attributes listed. The order in which participants rated the agents was randomised.

A questionnaire listing the same set of attributes was then given and participants asked to rate how important each one was for a given occupation. Again, the order in which participants rated the occupations was randomised.

Matching task

The matching task consisted of instructions informing participants under which criteria they were matching the agents and occupations (e.g. "Please match each occupation with the agent you think would be least appropriate in that role"). The list of eight occupations was listed horizontally across the page, with the images of the twelve agents displayed approximately 1.5" below. Participants were instructed to draw a line from each occupation to the agent they thought it matched.

Before completing the matching task, participants were shown an example using a different set of agents and occupations. This was done in order to demonstrate that

only one agent could be chosen for each occupation, but the same agent could be linked to more than one occupation. Different agents and occupations were chosen to avoid any biases in completing the actual questionnaire. The order in which the ratings and matching tasks were completed was counterbalanced to ensure that there were no order effects.

8.2.2 Results

Appropriate and Preference Data

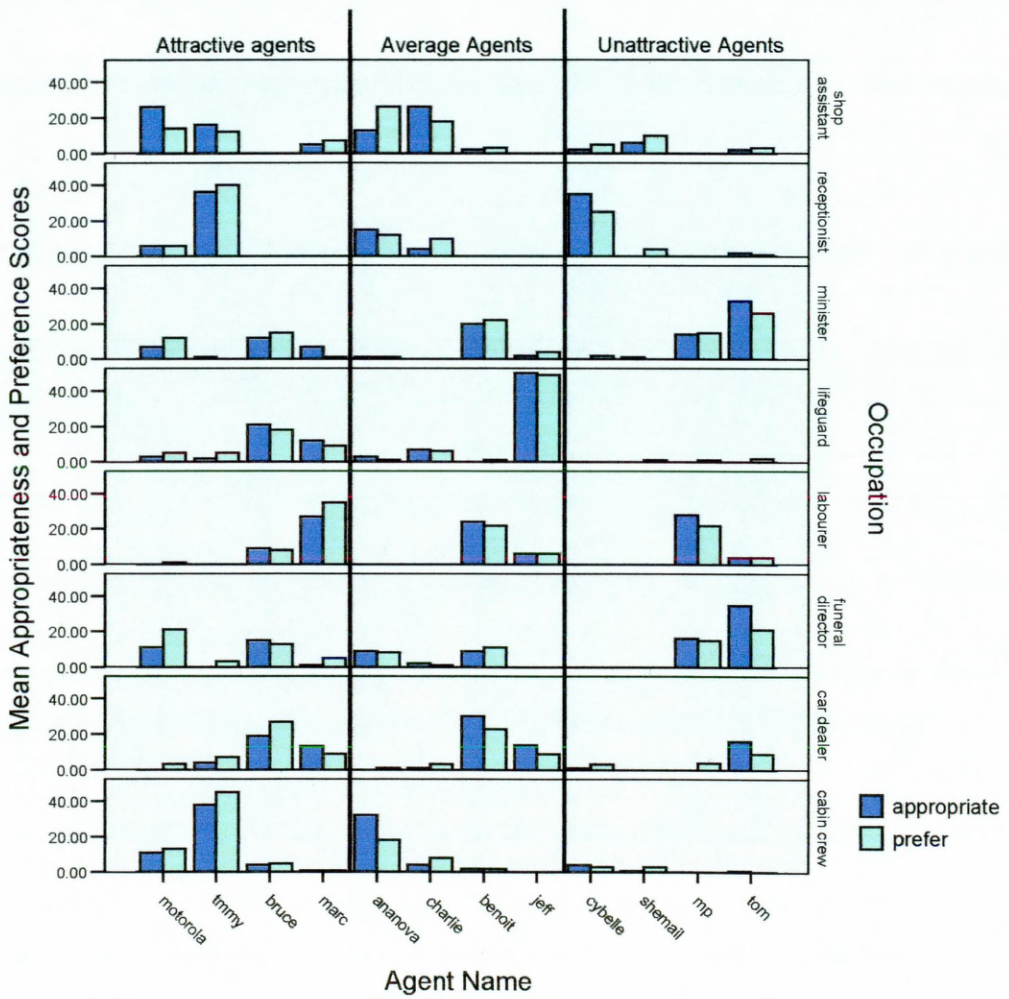


Figure 8.2: Comparison of appropriate and preference scores of agents for each occupation

Figure 8.2, above, shows that there is a relationship between appropriateness and preference scores, with, in general, agents that are considered appropriate for an

occupations are the same as those who are preferred. In addition, agents considered to be attractive or of average attractiveness score higher on appropriateness and preference when compared to unattractive agents. In general, attractive agents score highest on four out of the eight occupations (cabin crew, labourer, receptionist, and shop assistant). However, agents of average attractiveness only score highest for the occupation lifeguard, whereas unattractive agents score highest for funeral director and minister.

In order to test if appropriateness scores for occupations varied as a function of gender, eight separate Mann-Whitney U tests were carried out on each of the eight occupations and it was found that gender had a significant effect on six out of the eight occupation appropriateness scores (see Table 8.2, below). Female agents were considered the most appropriate for three out of the eight occupations (cabin crew, receptionist and shop assistant), and male agents were considered most appropriate for three out of the eight occupations (car dealer, labourer, and minister). There were no significant differences between appropriateness scores for male and female agents for funeral director or lifeguard. A similar patter of results was found for preference scores. Eight separate Mann-Whitney U tests were carried out and it was found that there were significant differences between male and female agents preference scores on six out of the eight occupations. Female agents were preferred significantly more than male agents for the role of cabin crew, receptionist and shop assistant, whereas male agents were considered preferred for the role of car dealer, labourer, and minister. There were no significant differences between preference scores for male and female agents for funeral director or lifeguard.

Table 8.2

The effect of gender on occupation appropriateness scores

Occupation		Mann-Whitney statistical summary [a]	Direction of result
Cabin Crew	A	U = 4.00, p = 0.023	Female
	P	U = 2.00, p = 0.010	Female
Car Dealer	A	U = 4.50, p = 0.027	Male
	P	U = 1.00, p = 0.006	Male
Funeral Director	A	U = 9.00, p = 0.142	n.s.
	P	U = 10.5, p = 0.226	n.s.
Labourer	A	U = 0.00, p = 0.002	Male
	P	U = 0.00, p = 0.003	Male
Lifeguard	A	U = 15.0, p = 0.618	n.s.
	P	U = 11.0, p = 0.253	n.s.
Minister	A	U = 1.50, p = 0.008	Male
	P	U = 3.50, p = 0.019	Male
Receptionist	A	U = 3.50, p = 0.013	Female
	P	U = 0.00, p = 0.003	Female
Shop Assistant	A	U = 2.00, p = 0.009	Female
	P	U = 1.00, p = 0.006	Female

Note: A – appropriateness results, P – preference result,

[a] N1=6, N2=6

In order to test if preference scores for occupations varied as a function of attractiveness, eight separate Kruskal Wallis tests were carried out and it was found that there was no significant effect of attractiveness on preference scores for any of the eight occupations (see Appendix 37 for statistical summary table). A similar result was found for appropriateness scores. Eight separate Kruskal Wallis tests were carried out and it was found that there was no significant effect of attractiveness on appropriateness scores for seven out of the eight occupations (see Appendix 38 for statistical summary table). However, there was a significant effect of attractiveness for the occupation of lifeguard (X^2 = 6.195, df = 2, p = 0.045). Follow up Mann-Whitney tests revealed that there was a significant difference between appropriateness scores for attractive and unattractive agents [U = 0.000, N1 = 4, N2 = 4, p = 0.014], with attractive agent being considered significantly more appropriate for the role of lifeguard when compared to unattractive agents. The differences between attractive agents and agent of average attractiveness, and unattractive agents and agents of average attractiveness were not significant.

In order to test the relationship between appropriateness and preference scores on each occupation, Pearson Product Moment Correlations were carried out. That is, for a given occupation, were the agents that were considered appropriate the same as those preferred? As shown in Table 8.3, results showed that there was a significant relationship between appropriate and preference scores, for all occupations.

Table 8.3

Correlation between appropriateness and preference for each occupation

Occupation	Relationship between Appropriateness and Preference
Cabin Crew	$r(10) = 0.946^*$, $p<0.001$
Car Dealer	$r(10) = 0.908^*$, $p<0.001$
Funeral Director	$r(10) = 0.925^*$, $p<0.001$
Labourer	$r(10) = 0.958^*$, $p<0.001$
Lifeguard	$r(10) = 0.882^*$, $p<0.001$
Minister	$r(10) = 0.859^*$, $p<0.001$
Receptionist	$r(10) = 0.937^*$, $p<0.001$
Shop Assistant	$r(10) = 0.950^*$, $p<0.001$

* Correlation significant at 0.01 level

In order to test the relationship between appropriate and preference scores, based on the attractiveness groupings of the agents, a Pearson Product Moment Correlation was carried out. It was found that there was a significant relationship between appropriate and preference for attractive agents [$r(29) = 0.928$], for unattractive agents [$r(29) = 0.966$], and for agents of average attractiveness [$r(29) = 0.934$]. In addition, there was a significant relationship between appropriateness and preference scores for male agents [$r(46) = 0.949$] and for female agents [$r(46) = 0.899$].

A Pearson Product Moment Correlation was also carried out to determine if there was a relationship between appropriateness and preference scores for each agent. As shown in Table 8.4, results showed that there was a significant relationship between appropriate and preference for eleven out of the twelve agents.

Table 8.4

Correlation between appropriateness and preference for each agent

Agent	Relationship between Appropriateness and Preference
Ananova 	r (8) = 0.755, p<0.05
Benoit 	r (8) = 0.976, p<0.001
Bruce 	r (8) = 0.931, p<0.001
Charlie 	r (8) = 0.892, p<0.05
Cybelle 	r (8) = 0.987, p<0.001
Jeff 	r (8) = 0.994, p<0.001
Marc 	r (8) = 0.925, p<0.001
Motorola 	r (8) = 0.685, p>0.005
MP 	r (8) = 0.979, p<0.001
Shemail 	r (8) = 0.889, p<0.05
Tmmy 	r (8) = 0.986, p<0.001
Tom 	r (8) = 0.977, p<0.001

Although there was a relatively high correlation between appropriate and preference scores for the agent "Motorola", this correlation was not significant. In this case there was an occupation for which she was considered appropriate for but not preferred and one for which she was preferred for but not considered appropriate for (see Figure 8.3). Motorola was considered to be the most appropriate agent for a shop assistant (along with "Charlie") but rated low on preference, and was rated low in appropriateness for a funeral director but was the agent that was most preferred (along with "Tom").

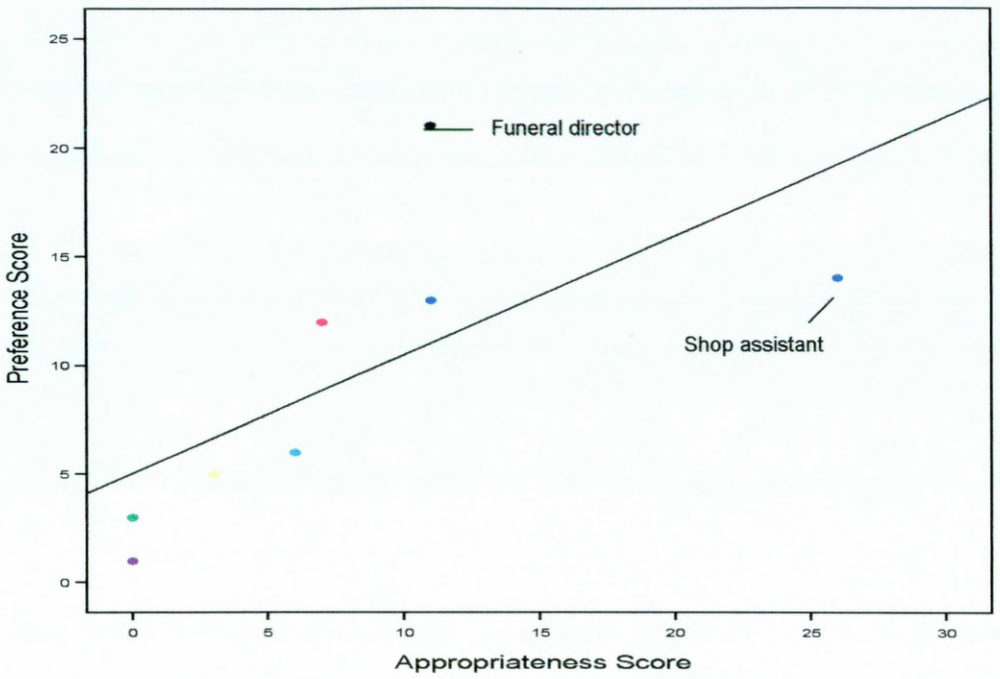


Figure 8.3: Relationship between appropriate and preference scores on each occupation for Motorola

In order to determine whether there was a relationship between occupation attributes and the attributes of the agents that were considered most appropriate for an occupation, separate Pearson Product Moment Correlations were performed for each occupation/agent pairing. Table 8.5 shows the occupations and the agents that were considered most appropriate in that role. Only one out of the eight occupations (car dealer – Benoit) had no linear relationship between the occupational attributes and the attributes of the agent considered most appropriate. Although the other seven occupations had significant correlations with the agent's attributes, these correlations were very weak.

Table 8.5

Relationship between agent and occupation attributes for the most appropriate agents

Occupation	Agent	Attributes	Pearson's Correlation[a]
Cabin Crew	Tmmy	Strength	r = 0.300, df = 96, p = 0.003
Car Dealer	Benoit	-	No significant correlations
Funeral Director	Tom	Affection	r = 0.200, p = 0.048
		Efficiency	r = -0.200, p = 0.049
		Attractiveness	r = 0.242, p = 0.016
		Clerical Skills	r = 0.341, p = 0.001
Labourer	MP	Social Skills	r = 0.202, p = 0.046
		Intelligence	r = 0.247, p = 0.003
		Efficiency	r = -0.230, p = 0.018
		Analytic Skills	r = 0.255, p = 0.011
Lifeguard	Jeff	Affection	r = 0.217, p = 0.032
		Efficiency	r = -0.204, p = 0.044
		Attractiveness	r = 0.208, p = 0.040
		Clerical Skills	r = 0.367, p = 0.000
Minister	Tom	Clerical skills	r = 0.318, p = 0.001
Receptionist	Tmmy	Intelligence	r = 0.342, p = 0.001
		Strength	r = 0.395, p = 0.000
		Spatial Skills	r = 0.246, p = 0.015
Shop Assistant	Charlie	Independence	r = 0.207, p = 0.041
		Strength	r = 0.214, p = 0.033
		Spatial Skills	r = 0.285, p = 0.004
	Motorola	Strength	r = 0.394, p = 0.000
		Spatial Skills	r = 0.260, p = 0.010

[a] df=96

Participant differences

In order to test for any gender differences between participants on the agent and occupation attribute rating scales Mann-Whitney analyses were carried out. It was found that there were no significant differences between male and female participants on 13 out of 17 agent attribute ratings (see Appendix 39 for statistical summary table). There were significant differences between male and female participants on 9 out of 17 occupation attribute ratings (see Appendix 40 for statistical summary table). However, correlational analysis of the mean ratings by male and female participants on each attribute revealed a significant, positive correlation ($r = 0.946$, $df = 15$, $p<0.001$), as shown in Figure 8.4 below.

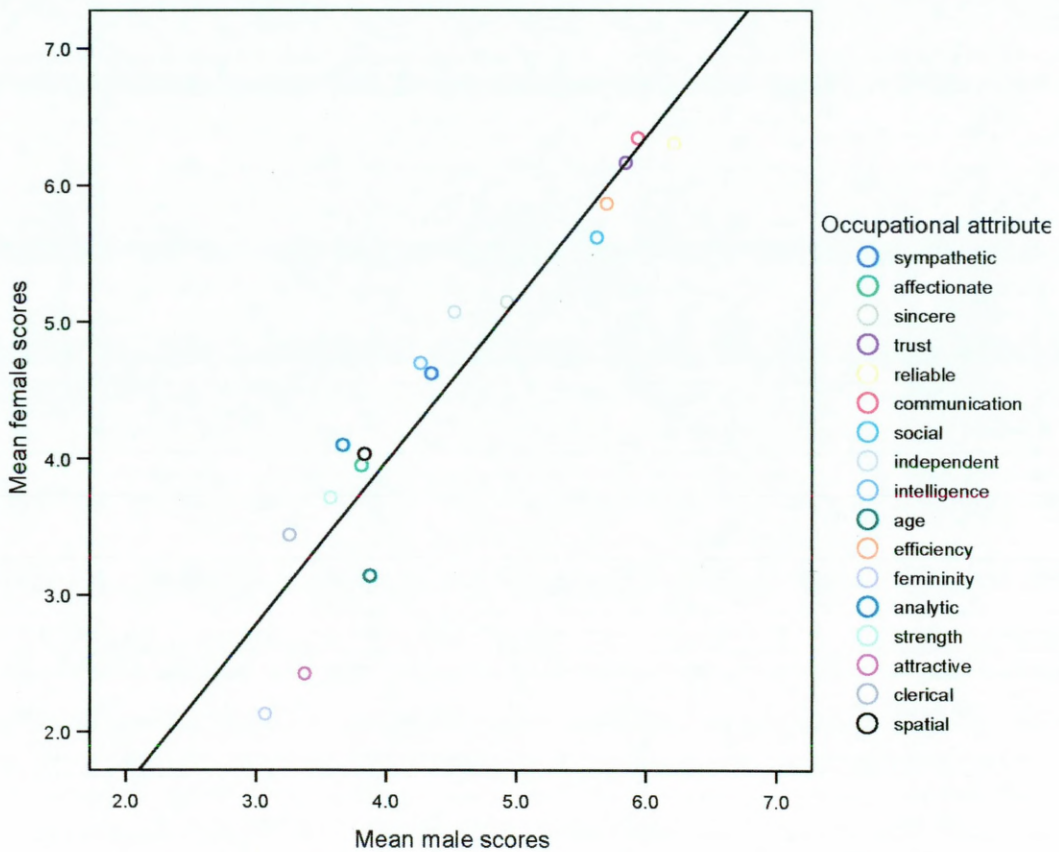


Figure 8.4: Correlation between male and female participants' means ratings on each occupational attribute

Therefore, it was decided to collapse across gender ratings and report overall ratings. The agent attributes on which male and female participants differed were sociability, independence, intelligence and strength, with female participants rating agents higher on all four attributes. The occupation attributes on which male and female

participants differed were trust, reliability, communication, independence, intelligence, age, femininity, analytic skill, and attractiveness. Female participants rated these attributes as being more important for the eight occupations than male participants did, with the exception of age, femininity and attractiveness.

Agent attribute differences

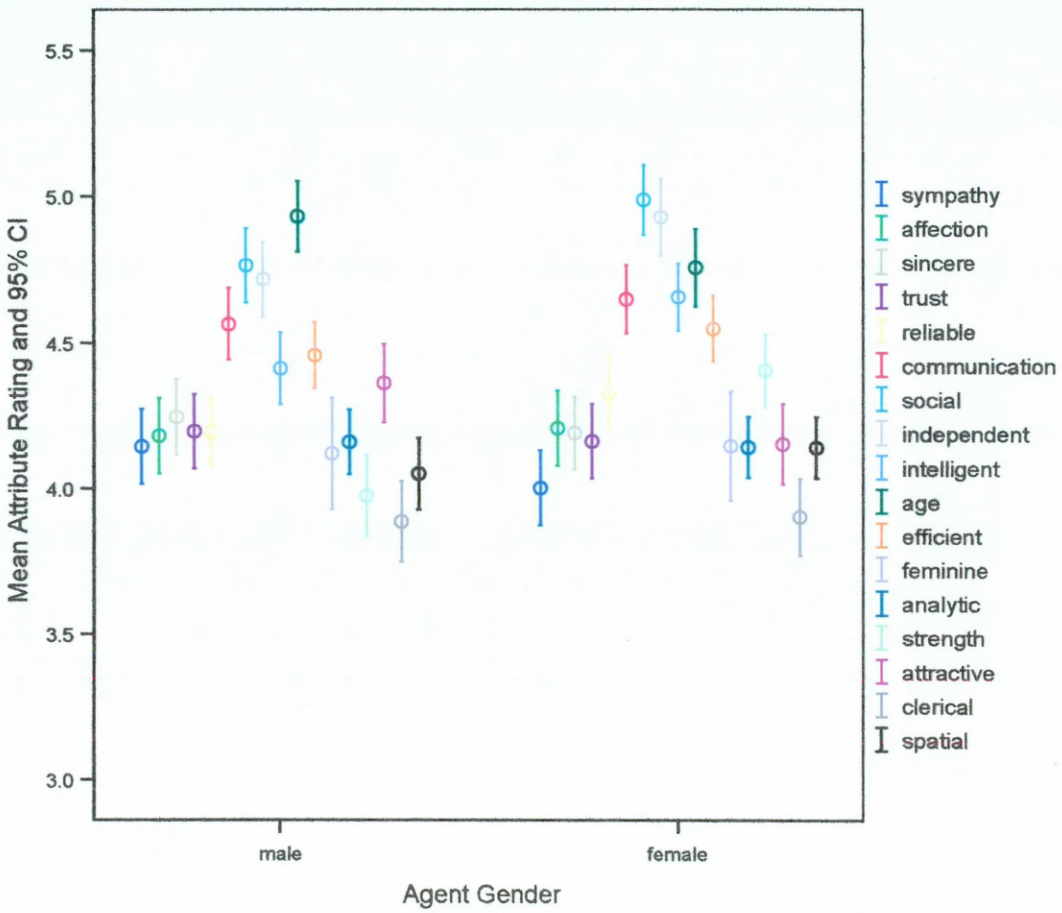


Figure 8.5: Mean Attribute Rating and 95% Confidence Intervals for each attribute, for male and female agents

In order to test the observed differences (see Figure 8.5) between male and female agents on the attributes, a Mann-Whitney analysis was carried out. It was found that there were significant differences between male and female agents on 16 out of the 17 attributes rated (see Appendix 41 for statistical summary table). Female agents were rated as being more sympathetic, affectionate, sincere, trustworthy, reliable, sociable,

intelligent, efficient, feminine, and attractive than male agents, as well being perceived to have better analytic, communication, and clerical skills, and as being younger than male agents. Male agents were rated as being more independent and having more strength than female agents. The only attribute on which male and female agents did not differ was spatial skills.

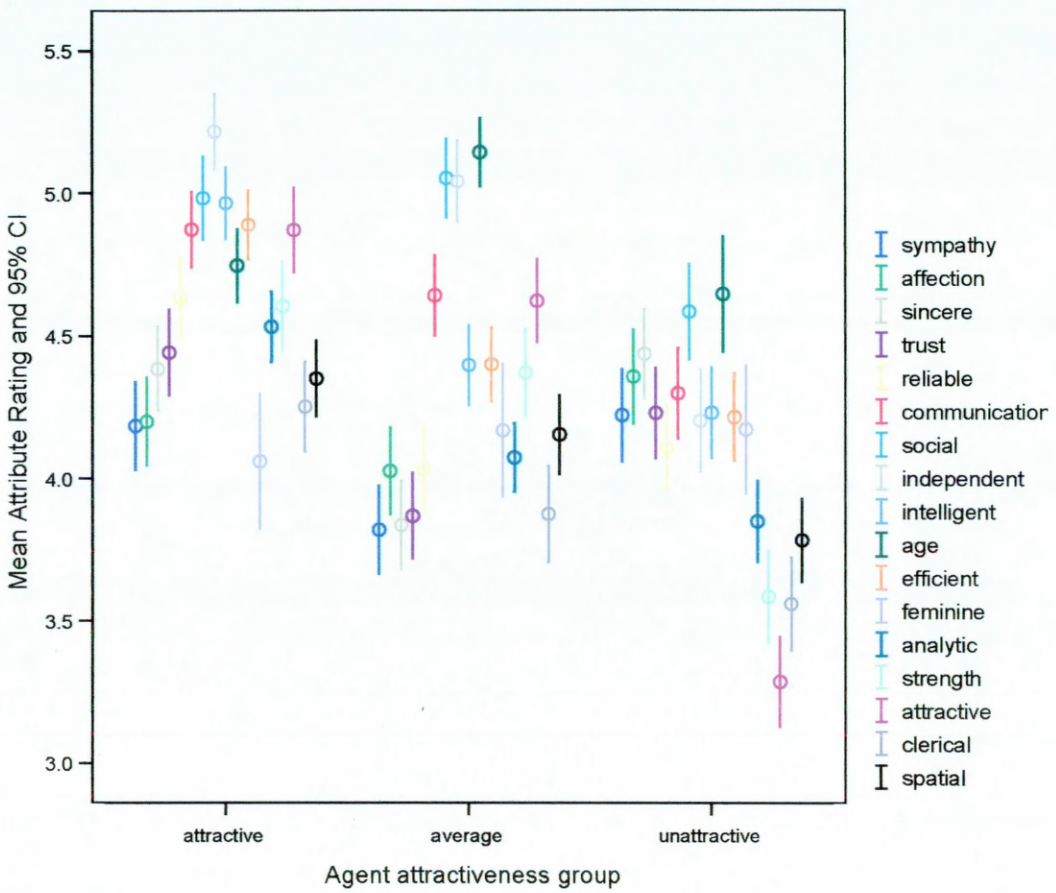


Figure 8.6: Mean Attribute Rating and 95% Confidence Intervals for each attribute, for attractive, average, and unattractive agents

In order to test the observed differences between attractive agents, unattractive agents, and agents of average attractiveness (see Figure 8.6) on the each of the attributes, a Kruskal Wallis test was carried out, and it was found that there was a significant effect of attractiveness for each of the seventeen attributes (see Appendix 42 for statistical summary table). Separate Mann-Whitney test were then carried out to investigate the difference between attractive and unattractive agents; attractive agents

and agents of average attractiveness; and unattractive agents and agents of average attractiveness (see Appendix 43 for statistical summary table).

Attractive agents were rated significantly higher on twelve out of the seventeen attributes when compared to agents of average attractiveness. These were sympathy, sincerity, trust, reliability, communication skills, intelligence, efficiency, analytic skills, strength, attractiveness, clerical skills, and spatial skills when compared to agents of average attractiveness. The only attribute on which agents of average attractiveness were rated significantly higher than attractive agents was age, where attractive agents were considered to be significantly older. Attractive agents were also rated significantly higher on eleven out of the seventeen attributes when compared to unattractive agents. These were reliability, communication skills, social skills, independence, intelligence, efficiency, analytic skills, strength, attractiveness, clerical skills, and spatial skills. Although there were significant differences between agents of average attractiveness and unattractive agents on eleven out of the seventeen attributes, unattractive agents were rated significantly higher on four of these (sympathy, affection, sincerity, and trust). Agents of average attractiveness were rated significantly higher on communication skills, social skills, age, strength, attractiveness, clerical skills, and spatial skills when compared to unattractive agents.

Agent attributes

A principal components factor analysis was conducted on the correlations of the seventeen agent attributes. Four factors were initially extracted with eigenvalues equal to or greater than 1.00. Orthogonal rotation of the factors yielded the factor structure given in Table 8.6. The first factor accounted for 24.8% of the variance, the second factor 22.2%, the third factor 13%, and the fourth factor 10.7%. The first factor seems to be work-related skills, the second factor empathy, the third factor personal characteristics, and the fourth factor male gender traits.

Table 8.6

Orthogonal factor loading matrix for seventeen agent attributes

Attribute	Factor 1	Factor 2	Factor 3	Factor 4
Intelligence	**0.834**	0.109	0.090	0.020
Efficiency	**0.814**	0.128	0.142	0.037
Analytic Skills	**0.811**	0.085	0.068	0.150
Independence	**0.662**	-0.056	0.148	-0.341
Spatial Skills	**0.625**	0.061	0.096	-0.184
Clerical Skills	**0.591**	0.144	0.135	0.511
Communication	**0.494**	0.389	0.471	-0.008
Sincerity	0.108	**0.863**	-0.015	0.121
Trustworthiness	0.251	**0.834**	0.033	0.122
Sympathy	0.035	**0.823**	0.220	0.092
Affection	-0.085	**0.777**	0.379	0.018
Reliability	0.581	**0.591**	0.009	0.069
Age	0.033	0.021	**0.749**	0.343
Attractiveness	0.345	0.088	**0.691**	0.109
Social Skills	0.095	0.415	**0.676**	-0.163
Strength	0.298	-0.106	0.035	**-0.799**
Femininity	0.158	0.152	0.345	**0.788**

Occupation Attributes

A second principal components factor analysis was conducted on the correlations of the seventeen occupation attributes. Six factors were initially extracted with eigenvalues equal to or greater than 1.00. Orthogonal rotation of the factors yielded the factor structure given in Table 8.7. The first factor accounted for 14.3% of the variance, the second factor 12.5%, the third factor 11.7%, the fourth factor 11.4%, the fifth factor 10.4%, and the sixth factor 9.2%. The first factor seems to be empathy, the second factor conscientious traits, the third factor methodical skills, the fourth factor communication skills (although this had a high negative loading of strength), the fifth factor physical appearance, and the sixth factor mentally strong/outgoing characteristics.

Table 8.7

Orthogonal factor loading matrix for seventeen occupation attributes

Attribute	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Sympathy	**0.905**	0.081	0.102	0.008	0.072	0.055
Affection	**0.876**	-0.062	0.080	0.145	0.049	0.122
Sincerity	**0.706**	0.381	0.208	-0.098	0.121	0.055
Reliability	0.160	**0.854**	-0.016	0.009	0.036	0.105
Trustworthiness	0.335	**0.771**	0.067	-0.117	-0.001	0.122
Efficiency	-0.235	**0.676**	-0.057	0.128	0.220	0.071
Strength	-0.078	0.153	**-0.811**	-0.005	0.074	0.220
Social Skills	0.192	0.078	**0.643**	0.163	0.020	0.395
Communication	0.289	0.224	**0.639**	0.047	0.209	0.308
Femininity	0.008	-0.050	0.122	**0.837**	0.075	-0.074
Attractiveness	-0.009	-0.050	0.138	**0.822**	0.188	-0.005
Age	0.094	0.164	-0.233	**0.645**	-0.232	0.151
Clerical Skills	0.062	0.051	0.298	0.018	**0.815**	-0.027
Spatial Skills	0.113	0.084	-0.404	0.138	**0.650**	0.135
Analytic Skills	0.076	0.150	-0.052	-0.002	**0.643**	0.396
Independence	0.018	0.068	-0.017	0.000	0.041	**0.842**
Intelligence	0.197	0.192	0.134	0.001	0.304	**0.620**

Table 8.9 shows the highest and lowest mean ratings on each attribute, and the corresponding agent (see Appendix 44 for all means). The most attractive agent was Tmmy, who was also considered to be the most feminine, trustworthy and have the best clerical and communication skills. Motorola was rated high on attractiveness, and was considered to be the most reliable, intelligent, efficient, and have the best analytic skills. In general, there was evidence of a positive general regard for female agents, as they were considered more attractive than male agents, and were rated highest on most attributes, with the only exceptions being independence, strength, social, and spatial skills. In addition, male agents were rated lowest on all attributes, with the exception of independence, strength, and spatial skills.

Table 8.9

Highest and Lowest Mean Rating on Each Agent Attribute (with standard deviations in parenthesis)

Attribute	Highest Rating	Lowest Rating
Affection	5.45 (1.30) Shemail	3.00 (1.37) Tom
Age	6.75 (0.70) Shemail	2.05 (1.15) MP
Analytic	5.11 (1.05) Motorola	3.12 (1.24) MP
Attractive	5.88 (1.06) Tmmy	1.85 (1.18) MP
Clerical	5.21 (1.24) Tmmy	2.51 (1.32) MP
Communication	5.48 (1.05) Tmmy	3.47 (1.50) MP
Efficient	5.43 (1.09) Motorola	3.88 (1.39) MP
Feminine	6.53 (0.89) Tmmy	1.76 (1.00) Bruce
Independent	5.49 (1.28) Bruce	2.78 (1.47) Shemail
Intelligent	5.65 (0.99) Motorola	3.23 (1.57) MP
Reliable	5.12 (1.30) Motorola	3.32 (1.56) Jeff
Sincerity	4.95 (1.93) Shemail	3.02 (1.53) Jeff
Social	5.66 (1.32) Jeff	3.82 (1.76) MP
Spatial	4.57 (1.35) Bruce	3.01 (1.30) Shemail
Strength	5.76 (1.11) Bruce	2.41 (1.25) Shemail
Sympathy	4.74 (1.59) Shemail	3.07 (1.27) Bruce
Trust	4.95 (1.36) Tmmy	3.18 (1.50) Tom

Table 8.10 shows highest and lowest mean ratings on each attribute, and the corresponding occupation (where the higher the score, the more important an attribute is). Minister was rated highly on five out of seventeen attributes, which are associated with empathy, such as sympathy, affection and trust. Lifeguard was also rated highly on five out of seventeen attributes, including reliability, efficiency, and spatial skills. Labourer was rated lowest on eleven out of seventeen attributes, but was rated high on strength.

Table 8.10

Mean Rating on Each Attribute for Each Occupation (with standard deviations in parenthesis)

Attribute	Highest Rating	Lowest Rating
Affection	5.87 (1.24) Minister	1.97 (1.26) Labourer
Age	4.30 (1.78) Lifeguard	2.96 (1.73) Shop Assistant
Analytic	4.25 (1.58) Receptionist	3.31 (1.60) Labourer
Attractive	4.37 (2.07) Cabin Crew	1.79 (1.32) Labourer
Clerical	5.83 (1.67) Receptionist	1.83 (1.16) Labourer
Communication	6.81 (0.41) Receptionist	4.44 (1.62) Labourer
Efficient	6.18 (1.21) Lifeguard	4.88 (1.64) Minister
Feminine	4.06 (2.12) Cabin Crew	1.78 (1.33) Labourer
Independent	5.34 (1.76) Lifeguard	4.44 (1.62) Funeral Director
Intelligent	5.10 (1.33) Minister	3.53 (1.52) Labourer
Reliable	6.69 (0.65) Lifeguard	6.01 (1.30) Car Dealer
Sincerity	6.58 (0.92) Minister	3.56 (1.81) Labourer
Social	6.33 (1.03) Cabin Crew	4.41 (1.38) Labourer
Spatial	4.74 (1.71) Lifeguard	3.00 (1.75) Minister
Strength	6.31 (1.11) Labourer	2.09 (1.38) Receptionist
Sympathy	6.65 (0.82) Minister	2.24 (1.38) Labourer
Trust	6.78 (0.78) Minister	5.55 (1.33) Labourer

Table 8.11

Most Important Attribute for Each Occupation

Occupation	Most Important Attribute
Cabin Crew	Communication Skills
Car Dealer	Communication Skills
Funeral Director	Reliability
Labourer	Strength
Lifeguard	Reliability
Minister	Trustworthiness
Receptionist	Communication Skills
Shop Assistant	Reliablity

Table 8.11 shows the most important attribute for each occupation. As can be seen, communication skills are perceived to be of high importance in service provision occupations such as cabin crew and receptionist.

8.2.3 Discussion

It was hypothesised that the appropriateness of an agent to an occupation would vary as a function of agent gender. This was partially supported, as gender of the agent had a significant effect on appropriateness scores for six out of the eight occupations (these were cabin crew, car dealer, labourer, minister, receptionist, and shop assistant). In addition, the direction of the results was congruent with the occupations (i.e. female agents were rated as significantly more appropriate for female dominated occupations, and male agents were rated significantly more appropriate for male dominated occupations). Although "shop assistant" was categorised in the male/low emotion category, it was still perceived as a "female" emotion, with female agents being rated significantly more appropriate for this role compared to male agents. Similarly, "minister" fell into the feminine/high emotion category, but was deemed to be significantly more appropriate for male agents compared to female agents. In addition, the same pattern of results was found for preference scores, with female

agents being preferred for female dominated occupations and male agents being preferred for male dominated occupations.

The only two occupations on which gender did not affect appropriateness or preference scores were funeral direction and lifeguard. However, for the occupation of lifeguard, 3 out of the 6 male agents were rated as not appropriate (i.e. they did not receive any "votes" for appropriateness). Thus, the standard deviation was large for the male agents. This could have accounted for the non-significant result for lifeguard, as, taking out the zero responses (for both male and female agents) showed that male agents were rated over four times more appropriate in the role of a lifeguard compared to female agents.

These results support the findings by Glick et al. (1995) who found that most occupations are strongly dominated by either males or females. They also support the idea of an association or "match-up" between the endorser and the product (McCracken, 1989; Till & Busler, 1998), in so far as male agents were not only considered as most appropriate for male dominated occupations, but they were also the most preferred in those roles. The same was true for female agents. This suggests that gender may be the most salient factor in determining the most appropriate and preferred agents for different types of roles.

However, it was also hypothesised that attractive agents would be most preferred over and above all other agents, particularly for service, sales, or public relation related occupations. Although the mean scores showed that attractive agents were considered most appropriate and were preferred over unattractive agents and agents of average attractiveness, particularly for sales or public relation occupations such cabin crew, receptionist, and shop assistant, and also for labourer, these results were not statistically significant. Mann-Whitney analysis revealed no significant differences between attractive, unattractive, and agents of average attractiveness on any of the occupations, for both appropriateness and preference scores. The only exception to this was for the occupation "lifeguard", where attractive agents were considered to be significantly more appropriate compared to unattractive agents. These findings contradict the finding by Heilman and Saruwatari (1979), who suggested that attractiveness is advantageous in sex congruent jobs, and that attractive people are

deemed to be "more qualified" for the job. They also found that unattractive females were rated more favourably when compared to attractive females, in a typically male role, whereas the results of the present study suggest that there is no effect of attractiveness on either male or female occupations.

In order to test for a relationship between the agents that were considered appropriate for a given occupation and those who were preferred, a Pearson Product Moment Correlation was carried out. It was found that there was a significant relationship between appropriate and preference scores for each occupation. It was also found that there was a significant relationship between appropriateness and preference scores for attractive agent, unattractive agents, and for agents of average attractiveness. In addition, there was a significant relationship between appropriateness and preference scores for male agents and for female agents. These results suggest that, irrespective of the attractiveness or gender of an agent, people prefer those agents who are most appropriate for a given occupation.

Correlational analysis was also carried out for each individual agent's appropriateness and preference scores, and was significant for all agents with the exception of "Motorola"  . Although there was a relatively high correlation between appropriate and preference scores for "Motorola", this correlation was not significant. In this case there was an occupation for which she was considered appropriate for but rated low on preference (shop assistant) and one for which she was preferred for but not considered appropriate for (funeral director). Although Motorola was considered most appropriate for a shop assistant (along with Charlie), Ananova was the agent who was most preferred. However, further analysis revealed that Motorola was rated as being significantly more sincere, trustworthy, reliable, intelligent, efficient, and analytic when compared to Ananova, thus these factors may influence the degree to which an agent is considered appropriate for a shop assistant. However, Ananova was the agent who was most preferred, and was rated significantly younger, and more feminine, when compared to Motorola, thus these attributes may be the most salient in terms of preference judgements for shop assistants. However, age and femininity were the attributes that were considered least important for a shop assistant. Perhaps the most salient aspect of a shop assistant was gender, as 97% of the appropriateness

decisions were for female agents. Although Motorola was the agent most preferred as a funeral director (along with Tom), Tom was considered to be the most appropriate. Again, gender may have been the most salient feature, as 83% of appropriateness judgements for a funeral director were for male agents. In general, however, the agents that were rated as most appropriate and preferred for an occupation were gender congruent with the occupation, which again suggests that gender may be the most salient factor when determining the most appropriate, and therefore the agent that will be most preferred, when choosing an agent to be used for a particular occupation.

Agent attribute differences

The results of the Mann-Whitney analysis revealed that there were significant differences between male and female agents on 16 out of the 17 attributes rated, and that there was, in general, a positive regard for female agents. Female agents were rated as being more significantly more sympathetic, affectionate, sincere, trustworthy, reliable, sociable, intelligent, efficient, feminine, and attractive when compared to male agents, as well being perceived to have better analytic, communication, and clerical skills, and as being younger than male agents. Male agents were rated as being significantly more independent and having more strength than female agents. The only attribute on which male and female agents did not differ was spatial skills.

Attractiveness (as a factor) was also found to have an effect on agent attributes. Attractive agents were rated significantly higher on twelve out of the seventeen attributes, including sincerity, trust, intelligence, attractiveness, and communication skills when compared to agents of average attractiveness. The only attribute on which agents of average attractiveness were rated significantly higher than attractive agents was age, where attractive agents were considered to be significantly older. Attractive agents were also rated significantly higher on eleven out of the seventeen attributes when compared to unattractive agents. Although there were significant differences between agents of average attractiveness and unattractive agents on eleven out of the seventeen attributes, unattractive agents were rated significantly higher on four of these (sympathy, affection, sincerity, and trust).

However, these results, overall, show that there is a general, positive regard for attractive agents when compared to unattractive agents and agents of average attractiveness, and for female agents compared to male agents. Thus, gender may be the most salient aspect when determining appropriateness and preferences, but the aesthetic appeal, or attractiveness, of an agent may also be an important attribute to consider when choosing an agent. These findings also partially support the literature that suggests that attractiveness has a major influence on perceptions of communication ability, likeability, and trustworthiness (Belch & Belch, 1998).

Consideration of individual agents revealed that the most attractive agent was Tmmy  , who was also considered to be the most feminine, trustworthy and have the best clerical and communication skills. Motorola  was also rated high on attractiveness, and was considered to be the most reliable, intelligent, efficient, and have the best analytic skills. Again, in general, there was evidence of a positive general regard for female agents, as they were considered more attractive than male agents, and were rated highest on most attributes, with the only exceptions being independence, strength, social, and spatial skills. In addition, male agents were rated lowest on all attributes, with the exception of independence, strength, and spatial skills.

Individual ratings on each attribute were also calculated for each occupation. These results showed 'Minister' was rated highest on five out of seventeen attributes: sympathy, affection, sincerity and trust, all of which are associated with empathy, and also intelligence. Lifeguard was also rated highly on five out of seventeen attributes, including reliability, efficiency, and spatial skills. Labourer was rated lowest on eleven out of seventeen attributes, but was rated high on strength. In addition, the most important attribute for each occupation was calculated, and showed that communication skills were perceived to be of high importance in the service provision occupations of cabin crew, car dealer and receptionist. Reliability was most important for the role of a funeral director, lifeguard, and shop assistant. The most important attribute for a minister to have was trust, and the most important for a labourer was strength.

In order to determine whether there was a relationship between occupation attributes and the attributes of the agents that were considered most appropriate for an occupation, separate Pearson Product Moment Correlations were performed for each occupation/agent pairing. Only one out of the eight occupations (car dealer – Benoit) had no linear relationship between the occupational attributes and the attributes of the agent considered most appropriate. Although the other seven occupations had significant correlations with the agent's attributes, these correlations were very weak (none of the correlation coefficients were above 0.4). These results suggest that an agent's perceived appropriateness to an occupation may be based on their attribute compatibility, but only to a certain degree. For example, an occupation requiring a high degree of trust, such as a minister, does not require the most appropriate agent (in this case, Tom) to have a similar rating of trust. That is, there was no correlation between Tom's trustworthiness score and the trustworthiness score of a minister. The only attribute on which there was a significant correlation was clerical skills.

In conclusion, these results suggest that although there is some degree of correlation between agent and occupation attributes, this may not be the most salient aspect when determining and agent's suitability to a role. From the results of this study, it can be suggested that the most salient aspect of an agent's suitability, or appropriateness, to an occupation is gender. Furthermore, gender also determines the agent participants would prefer for a particular occupation. Although the attractiveness of an agent influences perceptions made upon it, these perceptions do not drive either appropriateness or preference judgements.

Chapter 9 : The role of interaction and agent aesthetics on user perceptions

9.0 Aims of chapter

One of the ways in which interface agents are used is on the Internet to interact with visitors to a website. So far, only the images of agents have been considered. However, in order to try to understand which factors affect a user's perceptions of an agent, interaction must be taken into consideration.

This chapter aimed to examine the effects of interaction, and an interface agent's appearance on user's perceptions. A flight information task was designed to represent an opportunity to use an interface agent primarily as an alternative to looking up the information on a website. Participants were instructed to complete the task of finding out times and prices of a flight to New York, as well as information regarding the new visa regulations to the USA, via a Wizard of Oz set-up.

9.1 Introduction

Wizard of Oz (WoZ) studies are studies in which participants/users are given tasks, which involve interaction with a system, but they are unaware that the system is controlled/simulated by another person (the Wizard). One of the most common areas of research that WoZ methods are used in is the development of Natural Language Processing systems. By using a WoZ system, researchers are able to test functionality, user behaviour, and interfaces which are still at the design stage (i.e. have not been fully implemented). These findings can then be integrated into the design of the actual system.

An issue that has been suggested as important is that user perceptions of interface agents are mainly driven by the proficiency of the actual software system and the quality of interaction with the agent (Xiao et al., 2003), and that aspects such as the proficiency of the software, and the replies and suggestions made by the agent could strongly influence how much an agent is liked or disliked. While the influence of

varying levels of agent competence on user performance and perception could be examined, the main aim of this experiment was to examine if user's performance was enhanced, and if their perception of an agent was influenced, by agent attractiveness. Thus, agent competence was factored out as an influence. In addition, if user performance and subjective perception of an agent is negative (that is, participants have a general dislike for the agent), even though the system itself is competent, then this could be an important result for agent development. It could mean that there is a general dislike for agents, and agent technology, and this would have to be further investigated.

Maulsby, Greenberg and Mander (1993) suggest that the design of an agent system must be based on user's requirements and expectations. The conventional method of building an agent-based system can be expensive and may not always succeed in identifying these requirements and/or expectations. The actual systems are often so complex, that even small changes could result in major redesign. In addition, due to the features of agent systems, the agent application must be carefully constructed and debugged to ensure interaction is as smooth as possible.
Although Maulsby et al. warn one of the dangers in using a WoZ system is exaggerating the capabilities of the system by giving the Wizard unrealistic powers, which a real system could never have, this only has bearing in studies which are investigating actual interface or system design. Naturally, if system design were being investigated, then these exaggerations would cast doubt on the validity of the investigation and over-estimate results obtained. However, the current study is concerned with agent perception, and although system usability is being measured, it is not being used as a method for building a real, working agent system.

Salber and Coutaz (1993) define a Wizard of Oz system as one that "allows the observation of a user operating an apparently fully functioning system whose missing services are supplemented by a hidden wizard. The user is not aware of the presence of the wizard and is led to believe that the computer system is fully operational. The wizard observes the user through a dedicated computer system connected to the observed system over a network" (p. 95). Wizard of Oz techniques are often used when the technology for building the interface does not exist or is not sophisticated or advanced enough to be applied to 'real' applications (Dahlback, Jonsson and

Ahrenberg, 1993; Wilson and Rosenberg, 1988; Landauer, 1987). Some argue that the dialogue between a human and a computer should resemble the dialogue between humans. However, Dahlback and colleagues assert that human-human interaction cannot be used as a basis for the dialogue between a human and a computer. Some of the main differences between the two dialogues include complexity, interpersonal factors such as politeness, partners' background knowledge, and the use of a different communication channel (i.e. written keyboard input versus spoken dialogue). They suggest that the best way of developing software for use in a natural language environment is to study how users interact and converse with existing natural language interfaces (NLI). However, in order to develop the next generation of NLIs it would be ineffective to use existing systems. Thus, interaction data from Wizard of Oz techniques can be used, where users believe they are interacting with a computer system. However, they stress that Wizard of Oz simulations can be difficult to implement for a number of reasons, the main one being that there is much dissimilarity between humans and computers. The main differences are in consistency (people are adaptable, computers are rigid); speed of output (people much slower at typewriting, compared to computer output); and errors made (computers never make small mistakes, such as spelling errors). Consequently, in order to make users believe that they are interacting with a computer, the output produced (by the wizard) must resemble a computer, instead of a person typing the answers. Therefore, in order to simulate the timings and consistency of a computer, it is suggested that a set of pre-defined responses be used.

Gruen, Sinder, Boettner and Rich (1999) used both a Wizard of Oz simulation and interviews/observations on a system designed to help users with their e-mail. They found that participants, when told they were to treat the human experimenter as though they were interacting with a computer, used more low-level commands, whereas they expressed high-level goals and elicited a high-level task structure when told to treat the experimenter as a human helper. This highlights the differences in interaction style when interacting with a computer, and suggests that even when interacting with another person, if that person is 'pretending' to be a computer, users will treat them differently.

A Wizard of Oz methodology was employed in the current study to remove agent ability as a variable, although participants believed they were interacting with a fully functional interface agent. By using this approach the main focus can be on evaluating the perception of the interface agent, instead of the quality of a 'real' agent system. In some sense it provides an upper limit example of the performance that might one day be possible by an interface agent that could interact using natural language dialogue. By exaggerating the competence of an interface agent, that is, simulating a natural dialogue understanding, agent capability can be removed as a factor.

9.2 Current Study

In Chapter 8, results showed that agent appropriateness to an occupation was mainly driven by gender. In order to assess the true influence of agent attractiveness, therefore, it was decided that gender should be removed as a factor. Female agents were chosen for two reasons: Firstly because the roles provided by many interface agents are service provision (e.g. receptionist), and findings in Chapter 7 showed that these are stereotypical female occupations, and secondly because it was decided to use cabin crew as the occupation in the current experiment as it was thought that this would generate a substantial amount of interaction, and this was found to be a stereotypical female occupation.

In this experiment both performance and satisfaction dimensions were be measured in order to evaluate an interface agent.

The usefulness of an agent can be measured in different ways, depending on whether objective or subjective measures are required. At the objective end of the scale, dependent measures such as task performance can be measured in terms of time taken to complete the task. Time spent on an interaction task could constitute a measure of efficiency. However, a participant (or user) may spend more time using an agent simply because they like it or find it interesting (Ruttkay, Dormann & Noot, 2002) or because they found the system difficult to use (Buisine & Martin, 2003). Therefore, it was decided to impose a time limit on the task and carry out conversational analysis in order to assess task performance.

At the subjective end of the scale, a user's affective reaction to the agent can be measured. For example, factors such as how much the user liked the agent, how engaging they thought the agent was, how they perceived the agent in terms of it's personality, and how likely they would be to use the agent in the future could be measured. However, these subjective factors may not necessarily influence how effectively a user can complete a task; it may be that they can carry out a task regardless of how much they liked an agent. Conversely, these factors may predict the likelihood of long-term use of an agent.

The task was reasonably open (i.e. it took the form of a scenario). If the task had been complex, or of a more personal nature, then there would be danger of participants trying to finish the task as quickly as possible, thus not providing enough dialogue. Also, if participants are given too detailed instructions or examples in the task, there is a danger that the questions they asked would be biased, with participants attempting to duplicate the instructions or examples instead of doing things their own way. Conversely, if the task were too simple there would not be enough data to analyse. By taking the form of a scenario the task can allow a variety of outcomes, all of which must be considered as correct. It also allows there to be many different ways of reaching the outcome or goal of the task.

As the participants were deceived in this study, ethical approval was necessary. This deception was approved as participants were informed of it after the study was complete. Approval was also required in order to delay informing participants of the deception until after the whole study was complete. This was necessary because if participants found out about the nature of the study before it was finished, word could have spread to other participants and invalidated future data.

A between subjects design was employed in this study. The independent variable was the attractiveness of the agent, which had 2 levels – attractive and unattractive. In addition, a control condition was used, in which the participant interacted with the system but only used the chat dialogue box – there was no image of an agent shown.

The dependent variables were the attribute ratings of each agent, the scores on the system usability scale, pleasurability questionnaire, communication experience questionnaire, and the conversation during the task.

It was hypothesised that user perceptions of an interface agent would vary as a function of attractiveness of the agent. In addition, it was hypothesised that ratings of attributes would vary after interaction with an agent. Finally, it was hypothesised agent attractiveness would affect system usability, pleasureability of the interaction, and perceptions of conversation quality.

9.2.1 Method

9.2.1.1 Participants

38 participants took part in this study, 16 were male and 22 were female, and were randomly assigned to conditions. They were recruited from Abertay University by means of convenience sampling; however, it was decided to omit computing students from the sample as they may have been more likely to guess that the system was not of a true technical nature. Demographic information collected showed that computer experience backgrounds were balanced across conditions.

9.2.1.2 Materials and Apparatus

A consent form was given to each participant (see Appendix 45), which briefly described the nature of the experiment, as well as information regarding participant confidentiality/consent and contact details for any further questions.

An instruction sheet (see Appendix 46) described the interaction task. Participants were informed that they would be interacting with an interface agent, via text dialogue, who would be portraying the role of British Airways cabin crew, and their task was to find out times and prices of a flight to New York. Participants were also told that they could ask for any additional information about the flight. Finally, the

instructions stated that the interaction would be stopped after participants had completed the task or had been interacting with the agent for 10 minutes. After this time, participants were informed that the post (experiment) questionnaires should be completed.

There were five questionnaires used during this experiment. The first one was the agent attribute questionnaire (see Appendix 47), which was used pre- and post-experiment (except in the control condition when it was not used at all). At the top of the questionnaire participants were instructed to rate the agent shown on each of the dimensions listed. The image of the agent (either the attractive or unattractive agent) was displayed in colour, and was 1" x 1.5" in size. There were 12 bi-polar dimensions listed on a 1 – 5 scale (these included sympathy, intelligence, attractiveness, appeal, and usefulness). For each dimension, the "negative" end of the scales was represented by 1, and the "positive" end of the scale was represented by 5. Although, psychometrically, the scales should perhaps have been randomised so that they "switched" direction, previous experience has indicated that participants find this "consistent" scale easier to work with. Finally, participants were asked to rate their overall impression of the agent on a 1 – 5 scale, where 1 was bad and 5 was good.

There were four post-interaction questionnaires used. The first was the System Usability Scale (see Appendix 48), which was used to determine how easy participants found using the system, and was a slightly adapted version of the System Usability Scale developed by Brooke (1996). There were ten questions, each of which had to be answered using a 5-point likert scale, where 1 was "strongly disagree" and 5 was "strongly agree", and included questions such as "I thought the system was easy to use", "I thought there was too much inconsistency in this system", and "I felt very confident using this system". The questions were split so that half were positive statements and half were negative statements. The questionnaire was scored by firstly looking at the individual items. The score contribution of items 1, 3, 5, 7 and 9 was 1 minus the scale position. Similarly, the score contribution of items 2, 4, 6 and 8 was 5 minus the scale position. To obtain the overall value of system usability, the sum of the score contributions was multiplied by 2.5. The system usability scores ranged from 0 – 100.

The Communication Experience questionnaire (see Appendix 49) was used to assess the quality of communication, and was adapted from the questionnaire developed by Garau, Slater, Bee and Sasse (2001). There were fifteen questions, each of which had to be answered using a 5-point likert scale, with the anchors the same as described above. There were four sub-categories within this questionnaire: face-to-face, which measured how much the participant thought the interaction was like a real face-to-face conversation; involvement, which measured how involved participants felt in the conversation; co-presence, which measured the extent to which participants felt they were with, and interacting with, another person, rather than with a computer; and partner evaluation, which measured participants evaluation of their conversation partner, and how much they enjoyed the conversation. The questionnaire was scored by counting the number of positive responses in each sub-category (a positive response was one which was 4 or 5).

The Pleasurability questionnaire (see Appendix 50) was used to establish how pleasurable participants found the interaction, and was based on the questionnaire developed by Jordan (2000). There were ten questions, each of which had to be answered using a 5-point likert scale, with the anchors the same as described above, and included questions such as "I felt entertained when using this agent", "I enjoyed using this agent", and "This agent gave me a sense of satisfaction". The questionnaire was scored by summing the responses for each question.

The final questionnaire was concerned with participant's perceptions of the believability of the interaction (see Appendix 51), and included four questions such as "How believable was the agent", "Did you believe that the system was of a true technical nature", which were to be rated on a 5-point likert scale. The scale differed for each question. There were also ten open-ended questions, which the participants were free to answer as they wished, for example, "What did you like about the agent? What did you dislike?", and "Would you use the agent in the future".

In order to simulate a 'real' agent, the British Airways webpage (http://www.ba.com) was used as a template in order to represent the agent as cabin crew. The website was also used as a real source of information for the experimenter to use during the interaction task.

The two interface agents (see Figure 9.1) used in the experiment had been pre-selected from a sample of thirty-six female agents. 10 participants (5 males and 5 females) rated the agents on a 1-10 scale on attractiveness and age (where 1 was unattractive and 10 was attractive, and 1 was old and 10 was young). The attractive and unattractive agents selected were more than 1 standard deviation above and below the mean respectively. The two agents selected were also judged to be of a similar age, with neither of them falling 1 standard deviation above or below the mean age. The judgements of attractiveness of the agents were consistent across male and female participants.

Figure 9.1: Attractive and Unattractive Agents (l-r)

Participants were tested individually using a standard desktop PC with monitor, keyboard and mouse. The Wizard (experimenter) used a similar PC set-up. The two PC's were connected through a TCP/IP, with the participant's PC as the WebAgent server, and the experimenter's PC as the WebAgent client. Prior to the start of the experimental session, the experimenter started the WebAgent server application, which prompted for image and corresponding configuration file, and then started 'listening' for incoming connections (see Figure 9.2). The WebAgent client was then started by the experimenter and connected to the IP address/port of the participant's PC (see Figure 9.3). The WebAgent software was written in C#.

The interface agent was a coloured, static image, approximately 3cm x 3cm in size, and was placed at the upper-right hand corner of the webpage. As the webpage was a static image, none of the links were active and participants could not navigate around it as if it were a normal webpage (see Figure 9.4).

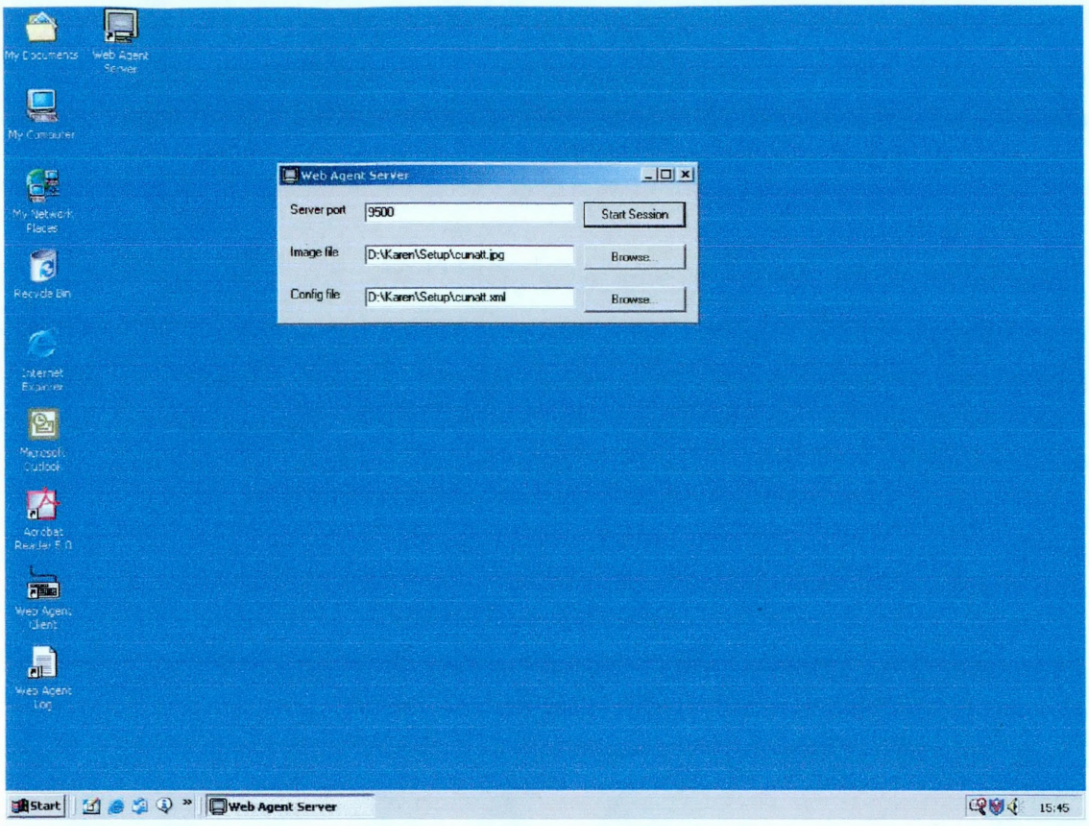


Figure 9.2: WebAgent Server

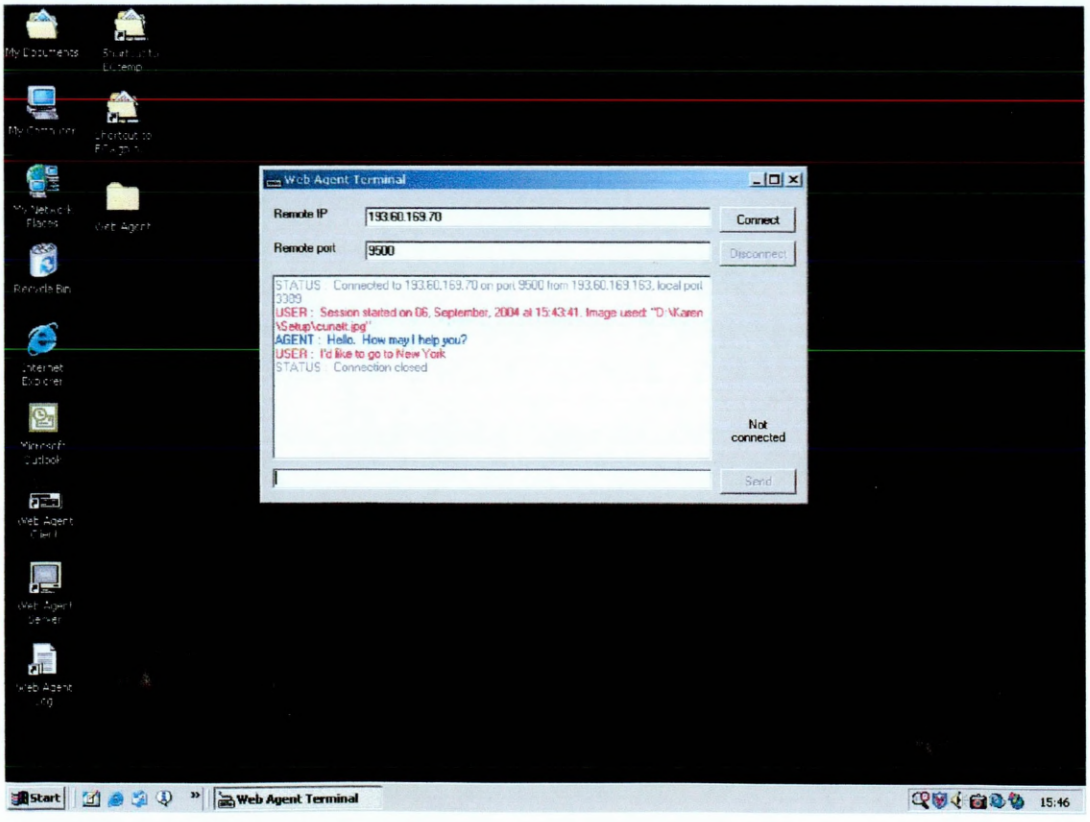


Figure 9.3: WebAgent Client

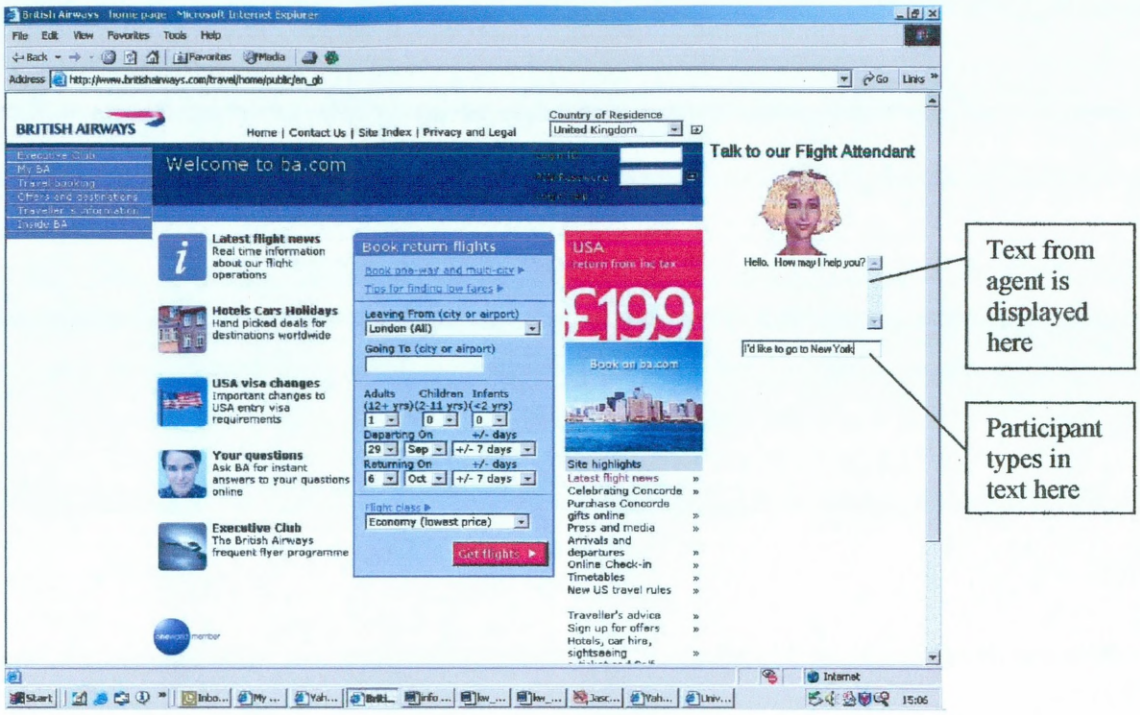


Figure 9.4: Web Interface

9.2.1.3 Procedure

Participants were asked to fill in the demographics/experience questionnaire which included 12 questions relating to computer experience and attitudes towards computers.

Pre-experiment questionnaire

Prior to commencing interaction, participants were shown the image of the agent on paper and asked to rate it on twelve attributes. These were sympathy, trustworthiness, communication skills, intelligence, age, femininity, attractiveness, memorability, appeal, friendliness, usefulness, how much they liked the agent, and their overall impression of the agent.

They were then given written instructions about the interaction task, and were shown the 'webpage' and instructed how to enter their dialogue with the agent.

Interaction task

As the agent was portraying the role of cabin crew, and to ensure interaction was similar across participants, the task required participants to find out times and prices of a flight to New York. Participants were also informed that they could ask any additional information about the actual flight (since they were 'talking' to a member of the cabin crew), such as information on the new visa regulations to the USA, in-flight entertainment, or meals. Pilot testing was conducted to ensure that the conversation would stimulate a sufficient amount of interaction, as well as to generate a number of pre-determined responses that could be used during the actual experiment.

As mentioned previously, the interface agent's responses were controlled through a Wizard of Oz technique. After explaining the experiment and the task to the participant, the experimenter left the room and went to an adjacent room where they monitored the dialogue with the participant. The participants entered text in a text box (0.5cm x 5cm in size), and then pressed the "Return" key. The 'agent' (Wizard) responded in a dialogue box above the text box. The experimenter monitored and responded to the dialogue using an editor window via a networked computer in an adjacent room. During the interaction, the dialogue exchanged between the participant and the wizard was simultaneously recorded in a log file. When participants had been interacting with the agent for 10 minutes, the experiment was concluded.

The 'wizard' determined how to answer the participant's questions, as well as asking any relevant questions (such as "which airport are you departing from"). A variety of responses were prepared for three main reasons; the first was to avoid typing/spelling mistakes, the second was to reduce the response time from the agent, and the third was to ensure consistency of the responses. However, it would have been impossible to provide guidelines or pre-defined responses for the Wizard instructing them how to respond in every possible situation. Therefore, the Wizard had to exercise good judgement and choose an appropriate response to questions that she was unprepared for.

In addition, to allow response times to be adequate for typing responses that were unprepared, all participants were informed that the system was quite slow as it was in the early stages of development.

Responses from the agents remained constant across conditions. For example:

Greeting:

Agent: Hello. How may I help you?

Prompt for information:

Agent: Which airport are you departing from?

The agent understood everything the user typed, just as a human partner would.

Post-task questionnaires and free-response interview

After completing the interaction task, participants were asked to complete the System Usability Scale, Communication Experience questionnaire, Pleasurability questionnaire, the agent attribute questionnaire, and the agent believability questionnaire. The System Usability Scale was used to determine how easy participants found using the system. The Communication Experience questionnaire was used to assess the quality of communication. The Pleasurability questionnaire was used to establish how pleasurable participants found the interaction. The agent attribute questionnaire was used to assess whether perceptions of the agents had changed after interaction. Finally, the agent believability questionnaire was used to determine how believable participants found the agent and the system. These questionnaires were completed in a random order.

9.2.2 Results

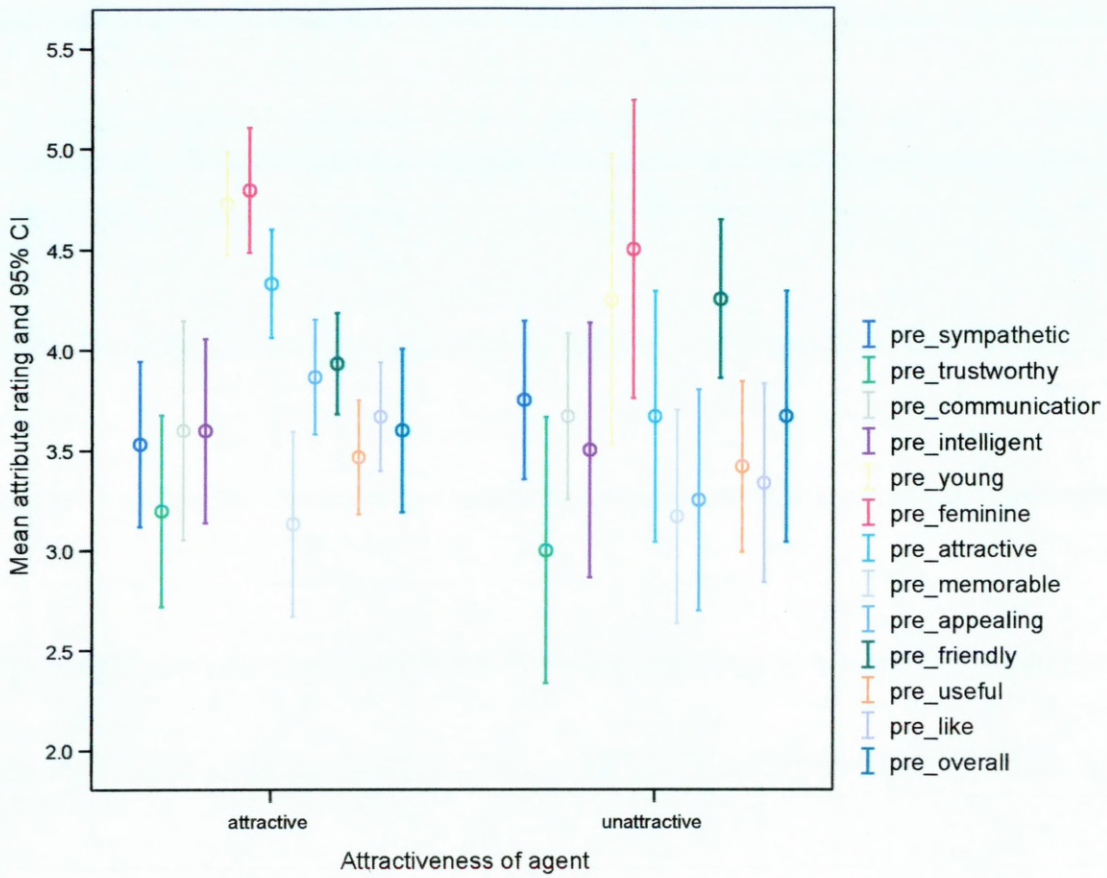


Figure 9.5: Mean pre-interaction attribute ratings for attractive and unattractive agents and 95% confidence interval

Figure 9.5, above, shows the pre-interaction mean scores on each attribute. For both attractive and unattractive agents, the attribute with the highest rating was femininity. The mean femininity score for the attractive agent was 4.80, with the 95% confidence interval falling between 4.48 and 5.11. The mean femininity score for the unattractive agent was 4.5, with the 95% confidence interval falling between 3.76 and 5.25. The lowest scoring attribute for the attractive agent was memorability, with a mean score of 3.13, and the 95% confidence interval falling between 2.67 and 3.60. The lowest scoring attribute for the unattractive agent was trust, with a mean score of 3.00, and the 95% confidence interval falling between 2.34 and 3.66.

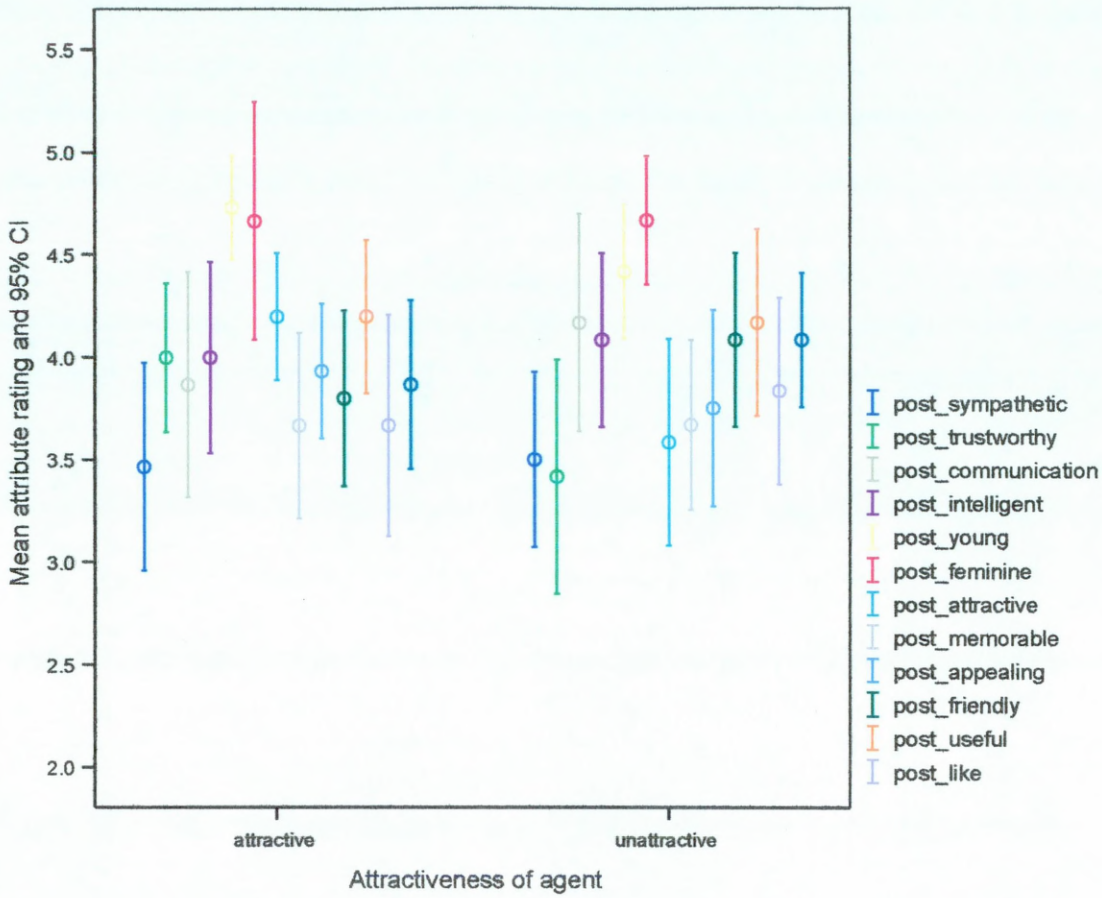


Figure 9.6: Mean post-interaction attribute ratings for attractive and unattractive agents and 95% confidence interval

Figure 9.6, above, shows the post-interaction mean scores on each attribute. For the attractive agent, the attribute with the highest rating was age. The mean age score for the attractive agent was 4.73, with the 95% confidence interval falling between 4.48 and 4.99. For the unattractive agent, the attribute with the highest rating was femininity. The mean femininity score for the unattractive agent was 4.67, with the 95% confidence interval falling between 4.67 and 4.98. The lowest scoring attribute for the attractive agent was sympathy, with a mean score of 3.47, and the 95% confidence interval falling between 2.96 and 3.97. The lowest scoring attribute for the unattractive agent was trust, with a mean score of 3.42, and the 95% confidence interval falling between 2.84 and 3.99.

In order to test the observed differences between attractive and unattractive agents on each of the attributes, both pre- and post-interaction, Mann Whitney U Tests were

carried out. It was found that there were very few significant differences between attractive and unattractive agents. The only attributes on which the agents differed were pre-interaction measures of appeal ($U = 55.5$, $N1 = 15$, $N2 = 12$, $p = 0.05$), and post-interaction measures of attractiveness ($U = 47.5$, $N1 = 15$, $N2 = 12$, $p = 0.025$). In both cases, the attractive agent was rated higher than the unattractive agent.

In order to test for pre-interaction and post-interaction differences for the attractive and unattractive agents, Wilcoxon Signed Ranks Tests were carried out. It was found that there were two attributes which differed pre- and post-interaction for the attractive agent, and three attributes which differed for the unattractive agent. For the attractive agent, post-interaction ratings of trustworthiness increased (Wilcoxon, $N = 8$, $z = -2.588$, $p = 0.010$), as did post-interaction ratings of usefulness (Wilcoxon, $N = 11$, $z = -3.317$, $p = 0.001$). For the unattractive agent, post-interaction ratings of appeal (Wilcoxon, $N = 5$, $z = -2.121$, $p = 0.034$), usefulness (Wilcoxon, $N = 8$, $z = -2.165$, $p = 0.030$), and likeability (Wilcoxon, $N = 6$, $z = -2.449$, $p = 0.014$) all increased compared to pre-interaction ratings.

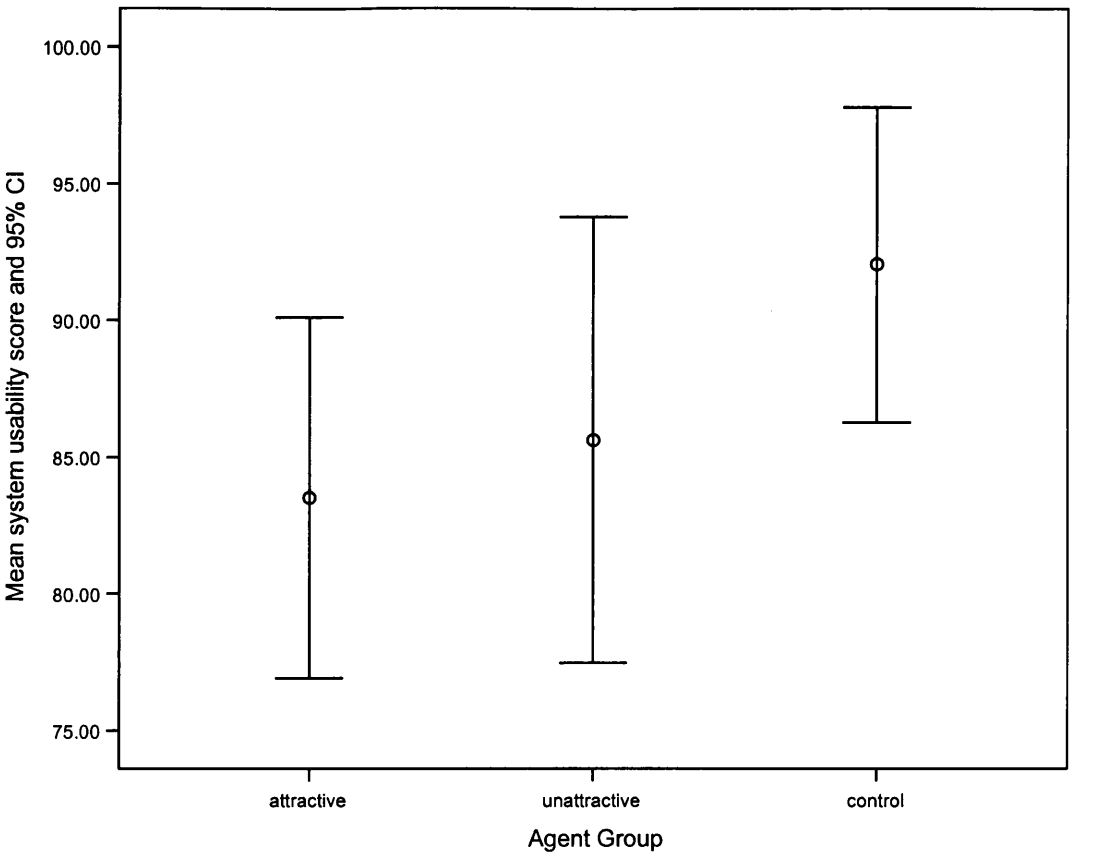


Figure 9.7: Mean system usability score and 95% confidence interval for each agent group

Figure 9.7 shows the mean system usability scores (SUS) for the attractive agent, unattractive agent and the control group (no picture). The mean SUS for the attractive agent was 83.5, with the 95% confidence interval falling between 76.90 and 90.09. The mean SUS for the unattractive agent was 85.62, with the 95% confidence interval falling between 77.46 and 93.78. The highest SUS was in the control group, with a mean of 92.04, and the 95% confidence interval falling between 86.28 and 97.80.

In order to test the observed differences between the attractive, unattractive and control groups on system usability scores, a one-way between subjects ANOVA was carried out and it was found that there was no significant effect of attractiveness group on system usability scores [$F(2,35) = 1.859$, $p = 0.171$], observed power = 0.361 (36%).

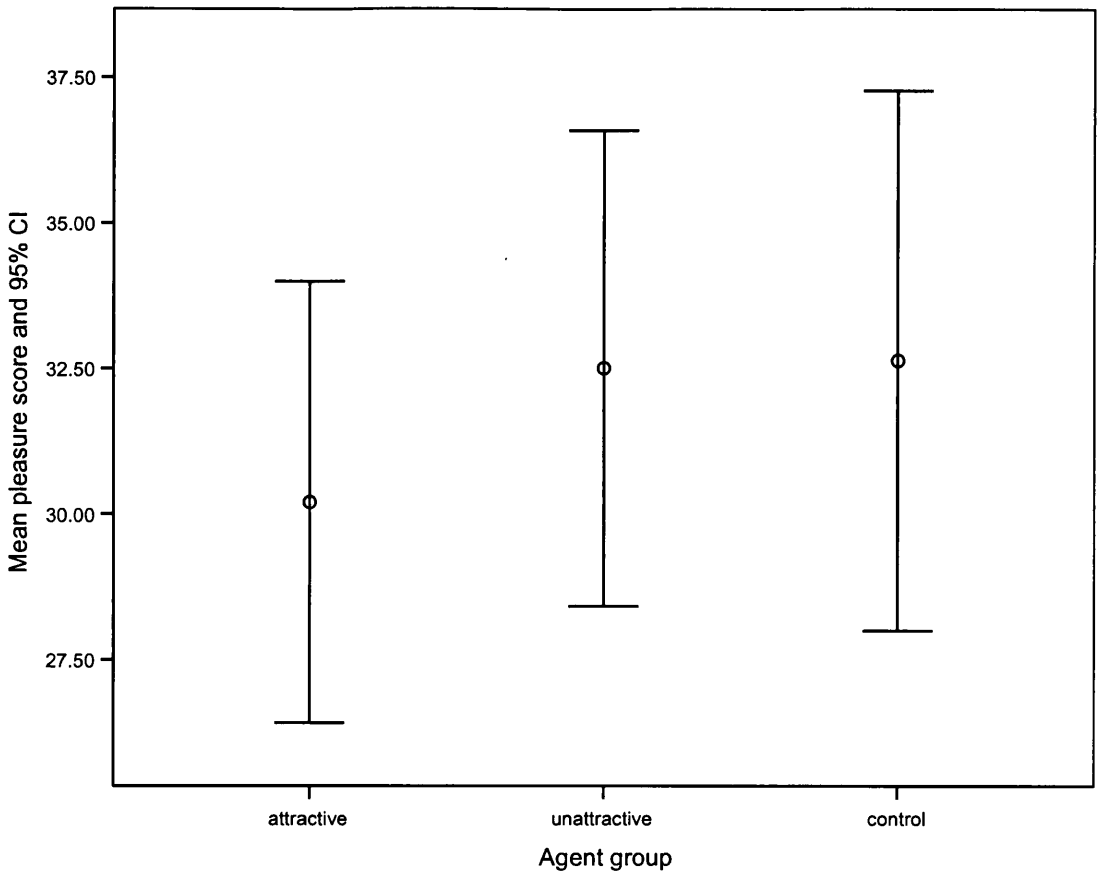


Figure 9.8: Mean pleasure score and 95% confidence interval for each agent group

Figure 9.8 shows the mean pleasure for the attractive agent, unattractive agent and the control group. The mean pleasure score for the attractive agent was 30.20, with the 95% confidence interval falling between 26.41 and 33.99. The pleasure score for the unattractive agent was 32.50, with the 95% confidence interval falling between 28.41 and 36.59. The highest pleasure score was in the control group, with a mean of 32.64, and the 95% confidence interval falling between 28.00 and 37.27.

In order to test the observed differences between the attractive, unattractive and control groups on pleasure scores, a one-way between subjects ANOVA was carried out and it was found that there was no significant effect of attractiveness group on pleasure scores [$F(2,35) = 0.562$, $p = 0.575$], observed power = 0.136 (13.6%).

The communication experience questionnaire was sub-divided into 4 categories: face-to-face; involvement; co-presence; and partner evaluation. Table 9.1, below, shows

the mean and standard deviation of the counts of 'high' responses (4 or 5) across the 6, 2, 2, and 5 questions in each condition, respectively.

Table 9.1

Mean and standard deviations of count responses of communication experience variables

Commuication experience	Attractive Agent	Unattractive Agent	Control
Face-to-face	3.53 (1.46)	4.08 (1.56)	4.27 (1.49)
n = 6			
Involvement	0.87 (0.64)	1.25 (0.62)	1.27 (0.79)
n = 2			
Co-presence	0.67 (0.82)	0.83 (0.83)	1.36 (0.67)
n = 2			
Partner evaluation	2.00 (1.36)	2.25 (1.54)	2.27 (1.00)
n = 5			

n = number of questions in category

The results in Table 9.1 show that the control condition, in which there was no picture of an agent present, produced the highest responses across all communication categories, followed by the unattractive agent, with the attractive agent scoring lowest across all communication categories.

Conversational Analysis

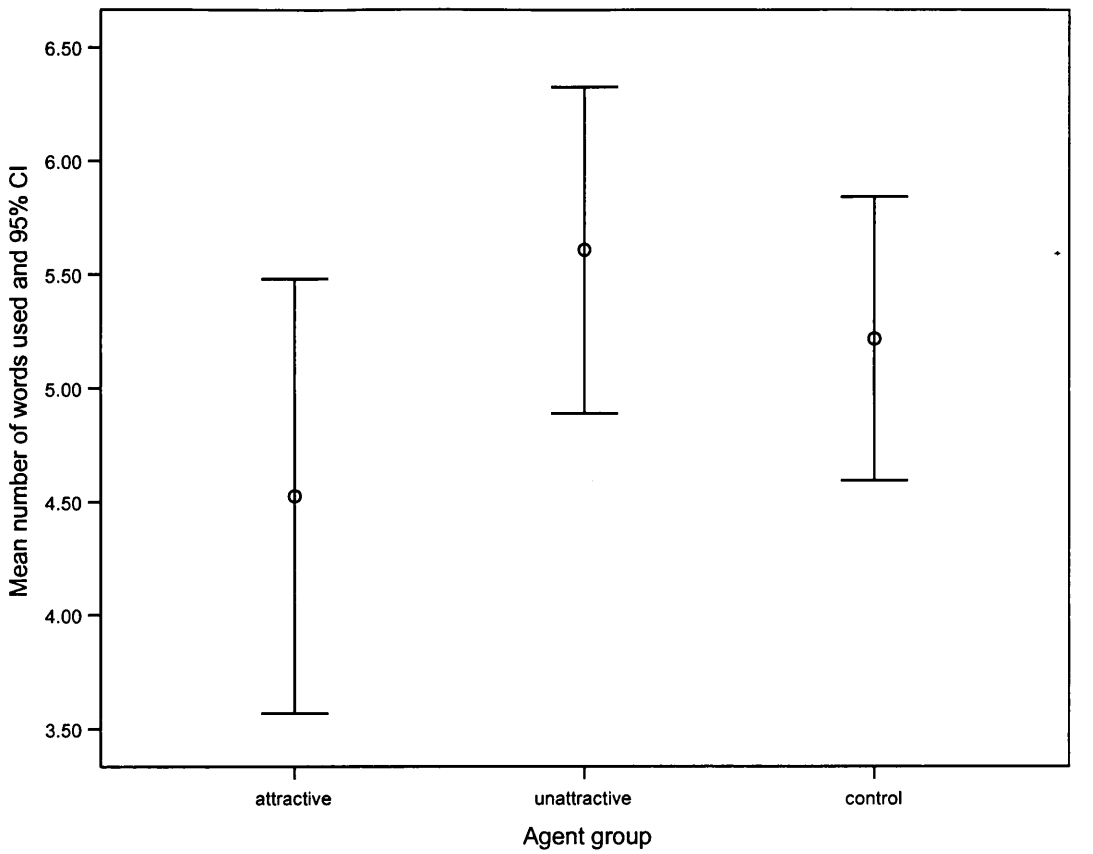


Figure 9.9: Mean number of words used during interaction for each agent group

Figure 9.9 shows the mean number of words used in each agent group during the interaction. The mean number of words used when interacting with the attractive agent was 4.52, with the 95% confidence interval falling between 3.57 and 5.48. The mean number of words used when interacting with the unattractive agent was 5.61, with the 95% confidence interval falling between 4.89 and 6.33. The mean number of words used in the control condition was 5.21, with the 95% confidence interval falling between 4.60 and 5.84. The greatest number of words used during the interaction was with the unattractive agent, followed by interacting with no picture of an agent present, with the lowest number of words being used in the interaction with the attractive agent.

In order to test the observed differences between the attractive, unattractive and control groups on the average number of words used, a one-way between subjects ANOVA was carried out and it was found that there was no significant effect of

attractiveness group on the average number of words used [F(2,35) = 2.220, p = 0.124], observed power = 0.422 (42.2%).

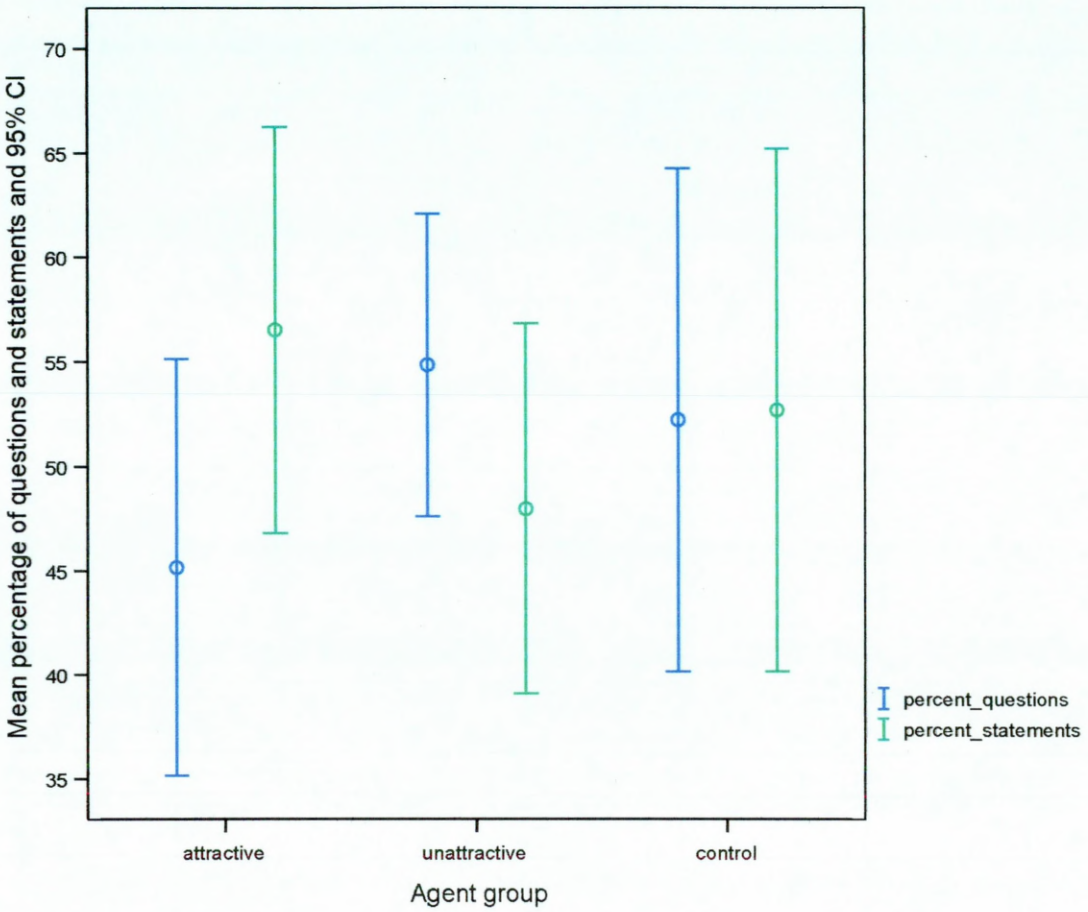


Figure 9.10: Mean percentage of questions asked and statements used during interaction for each agent group

Figure 9.10 shows the mean percentage of questions asked and statements used during interaction for each agent group. The mean percentage of questions asked when interacting with the attractive agent was 45.18, with the 95% confidence interval falling between 35.20 and 55.16. The mean percentage of questions asked when interacting with the unattractive agent was 54.87, with the 95% confidence interval falling between 47.63 and 62.11. The mean percentage of questions asked in the control condition was 52.22, with the 95% confidence interval falling between 40.17

and 64.27. The highest number of questions asked during the interaction was with the unattractive agent, followed by interacting in the control condition, with the lowest number of questions being asked during the interaction with the attractive agent.

In order to test the observed differences between the attractive, unattractive and control groups on the mean percentage of questions asked, a one-way between subjects ANOVA was carried out and it was found that there was no significant effect of attractiveness group on the mean percentage of questions asked [$F(2,35) = 1.303$, $p = 0.285$], observed power = 0.263 (26.3%).

The mean percentage of statements used when interacting with the attractive agent was 56.55, with the 95% confidence interval falling between 46.82 and 66.28. The mean percentage of statements used when interacting with the unattractive agent was 47.98, with the 95% confidence interval falling between 39.11 and 56.85. The mean percentage of statements used in the control condition was 52.68, with the 95% confidence interval falling between 40.16 and 65.20. The highest number of statements used during the interaction was with the attractive agent, followed by interacting in the control condition, with the lowest number of statements used during the interaction with the unattractive agent.

In order to test the observed differences between the attractive, unattractive and control groups on the mean percentage of statements used, a one-way between subjects ANOVA was carried out and it was found that there was no significant effect of attractiveness group on the mean percentage of statements used [$F(2,35) = 0.862$, $p = 0.431$], observed power = 0.186 (18.6%).

In order to test if there was a significant difference between the percentage of questions asked and the percentage of statements used, for each agent group, paired samples t-tests were carried out. It was found that there was no significant between percentage of questions asked and the percentage of statements used for the attractive agent [[$t(14) = -1.250$, $p = 0.232$], nor for the unattractive agent [$t(11) = 0.953$, $p = 0.361$], or the control group [$t(10) = -0.044$, $p = 0.966$].

Believability of the system

In order to test if believability of the system was an influencing factor in judgements made about system usability and pleasure, those participants with scores of 4 or 5 (where 5 indicated that they "definitely" thought they were talking to a human) were excluded, and two one-way ANOVAs were carried out to investigate for differences between the attractive, unattractive and control groups on system usability scores and pleasure scores. It was found that there was no significant effect of attractiveness group on system usability scores [$F(2,25) = 2.087$, $p = 0.145$], observed power = 0.388 (38.8%) nor on pleasure scores [$F(2,25) = 0.619$, $p = 0.546$], observed power = 0.142 (14.2%).

Further analysis was carried out to investigate if there were significant differences between those participants who thought they were talking to a human (score = 4 or 5; N = 10), and those who did not think they were talking to a human (score = 1, 2, or 3; N = 28). Two independent samples t-tests were carried out and it was found that there was no significant effect of "believability" on system usability scores [$t(36) = -0.267$, $p = 0.791$], nor on pleasure scores [$t(36) = -1.684$, $p = 0.101$].

9.2.3 Discussion

It was hypothesised that user perceptions of an interface agent would vary as a function of attractiveness of the agent. Mann Whitney U tests were carried out in order to test the observed differences between the attractive and unattractive agents on each of the attributes, both pre- and post-interaction. It was found that there were very few significant differences between attractive and unattractive agents. The only attributes on which the agents differed were pre-interaction measures of appeal, and post-interaction measures of attractiveness. In both cases, the attractive agent was rated higher than the unattractive agent.

In addition, it was hypothesised that ratings of attributes would vary after interaction with an agent. The pre-interaction mean scores showed that for both the attractive and unattractive agents, the attribute with the highest rating was femininity. The

lowest scoring attribute for the attractive agent was memorability, and the lowest scoring attribute for the unattractive agent was trust. The post-interaction mean scores showed that for the attractive agent, the attribute with the highest rating was age. For the unattractive agent, the attribute with the highest rating was femininity. The lowest scoring attribute for the attractive agent was sympathy, and the lowest scoring attribute for the unattractive agent was trust. In order to test for pre-interaction and post-interaction differences for attractive and unattractive agents, Wilcoxon Signed Ranks Tests were carried out. It was found that there were two attributes which differed pre- and post-interaction for the attractive agent, and three attributes which differed for the unattractive agent. For the attractive agent, post-interaction ratings of trustworthiness and usefulness increased. For the unattractive agent, post-interaction ratings of appeal, usefulness, and likeability all increased compared to pre-interaction ratings.

Finally, it was hypothesised agent attractiveness would affect system usability, pleasureability of the interaction, and perceptions of conversation quality. Although the mean scores revealed that the control group (no picture of an agent present) scored highest in terms of system usability and pleasure when using the agent/system, Analysis of Variance showed that there was no significant effect of attractiveness group on system usability scores, or on pleasure scores. The communication experience questionnaire was sub-divided into 4 categories: face-to-face; involvement; co-presence; and partner evaluation. Again, the results of this show that the control condition produced the highest responses across all communication categories, followed by the unattractive agent, with the attractive agent scoring lowest across all communication categories.

As there were no significant effects found for any of the measures taken, it was decided to investigate the interaction itself, that is, the 'conversation' between the participant and the agent. The mean number of words was calculated (by dividing the number of words used by the amount of times the participant 'spoke'), as were the mean percentage of questions asked (e.g. "what are the visa requirements in USA?), and the mean percentage of statements made (e.g. "Edinburgh"). Results showed that most words were used during the interaction with the unattractive agent, followed by interacting with no picture of an agent present, with the lowest number of words being

used in the interaction with the attractive agent. However, ANOVA revealed that that there was no significant effect of attractiveness group on the average number of words used.

Similarly, the highest percentage of questions asked during the interaction was with the unattractive agent, followed by interacting with no agent present, with the lowest percentage of questions being asked during the interaction with the attractive agent. However, ANOVA revealed that there was no significant effect of attractiveness group on the mean percentage of questions asked. The highest percentage of statements used during the interaction was with the attractive agent, followed by interacting with no agent present, with the lowest percentage of statements used during the interaction with the unattractive agent. Again, ANOVA revealed that there was no significant effect of attractiveness group on the mean percentage of statements used.

Further analysis was carried out to test for a significant difference between the percentage of questions asked and the percentage of statements used, for each agent group. Paired samples t-tests were carried out, and it was found that there was no significant between percentage of questions asked and the percentage of statements used for the attractive agent, nor for the unattractive agent, or the control group.

Finally, in the post interaction questionnaire one question asked about the believability of the system, or more specifically, if the participant thought that they were talking to a human. In order to test if this was an influencing factor in judgements made about system usability and pleasure, those participants with scores of 4 or 5 (where 5 indicated that they "definitely" thought they were talking to a human) were excluded, and two one-way ANOVAs were carried out to investigate for differences between the attractive, unattractive and control groups on system usability scores and pleasure scores. It was found that there was no significant effect of attractiveness group on system usability scores, or on pleasure scores.

Further analysis was carried out to investigate if there were significant differences between those participants who thought they were talking to a human (score = 4 or 5), and those who did not think they were talking to a human (score = 1, 2, or 3). Two

independent samples t-tests were carried out and it was found that there was no significant effect of "believability" on system usability scores, or on pleasure scores between the two groups.

Overall, these findings suggest that agent attractiveness, or, in fact, the presence, or absence, of an agent, may not impact on judgements made about the agent, or the system. In general, the attractive agent and unattractive agent were rated as being comparatively similar across all attributes measured. The only exceptions were on pre-interaction measures of appeal and post-interaction measures of attractiveness, on which the attractive agent was rated more favourably. Although there were few attributes which differed pre- and post-interaction for both attractive and unattractive agents, it is interesting to note that both attractive and unattractive agents were perceived to be more useful post-interaction. The attractive agent was also considered to be more trustworthy, post-interaction, and the unattractive agent was considered to be more appealing and likeable. These results suggest that perhaps the most salient aspect in agent perception is the actual interaction, with perceptions of the usefulness/usability of the agent increasing significantly after interaction.

In addition, it was found that agent attractiveness (and presence) did not affect System Usability Scores. That is, the system was perceived to be equally easy to use, regardless of the attractiveness of the agent, or if there was an agent present at all. In fact, the mean scores, although not significant, suggested that participants in the control condition, where there was no physical presence of an agent, found the system easier to use compared to when there was an agent present. The pleasureability scores also follow the same pattern of results, although, again, these results were not significant. Participants in the control condition found the interaction more pleasurable compared to when there was an agent present. In addition, interaction with an unattractive agent was found to be more pleasurable when compared to an attractive agent. However, power analysis of these results revealed very low observed power (36% for system usability and 14% for pleasureability), therefore these results are inconclusive. It remains possible that the research hypothesis could be false, and that there was no effect of agent attractiveness, however because the study had low power (e.g. there were too few participants), more empirical work would be required to rule this possibility in or out.

Nevertheless, the same pattern of results was found with respect to communication experience; the agent in the control condition higher was rated higher across all four sub-categories of communication experience (face-to-face, involvement, co-presence, and partner evaluation). This result is interesting as it suggests that participants found communicating with an agent more engaging when there was no physical presence of the agent, compared to when there was a picture of an agent (although, again, this was not a significant finding).

However, the images of the agents used were static, and this may have been an influencing factor. It may be the case that the agents would have to be more 'dynamic' in order for them to be considered more engaging. As the findings in this experiment were not significant, it may simply have been the case that the images of the agents were 'ignored' and participants concentrated on the interaction. This idea is further supported with the findings from the conversational analysis, which showed that there were no significant differences between the attractive agent, unattractive agent and control group in terms of conversational content. These findings are consistent with those of Xiao et al. (2003) who found that interaction quality and software proficiency were the most important aspects in user perceptions of interface agents. Although agent competence was not measured in the current experiment, the proficiency of the software/agent was consistent across the agent categories (attractive, unattractive, and control) and therefore could be an explanation as to why these agents were rated similarly across the measures taken.

Chapter 10 : General discussion and conclusions

10.1 Aims of the thesis

One key concept central to this thesis was the reported "beauty is good" stereotype (e.g. Dion, Berscheid & Walster, 1972; Eagly, Ashmore, Makhijani & Longo, 1991). These findings suggested that fundamental differences exist in the way(s) that people interact and judge others. It was noted that very few human-computer interaction studies have reported differences between attractive and unattractive agents and users' perceptions of them, and therefore the main aim of this thesis was to investigate the effect of aesthetics on agents. The investigation of these aesthetic differences was carried out over a series of experiments, which were designed according to the findings of each preceding experiment in order to investigate the potential influence of aesthetics, gender, occupation, context and, subsequently, interaction. The purpose of this chapter is therefore to outline the main findings of each experimental study carried out as part of this work and highlight any questions or points for discussion that may result from these. The main implications of these findings will be discussed alongside limitations of the studies, and future studies.

10.2 Main findings within each chapter

The experiments carried out reveal several important points and these will be discussed in the following sections.

10.2.1 Chapter 4

In the first experimental study, participants were informed that they were rating agents to be used by a financial institution on their website. The results showed that agents were differentiated by two factors: 'human-ness' and gender. 'Human-ness' related to whether the agent was typically human looking (i.e. male or female), or cartoon looking (i.e. a dog, or a frog). Gender related to the 'sex' of the agent, and had female agents at one end of the scale and male agents at the other end, with

cartoon agents in-between. Analysis of the preference data revealed that female agents were most preferred, followed by cartoon agents, with male agents being least preferred. This preference for female agents was also reflected in the ratings of the attributes being measured, with two of the female agents being rated highest on all nine attributes. Although factor analysis revealed that there were two main dimensions within the attributes (aesthetic and usability), the overall impression of an agent was determined by ratings of appeal and beauty.

The concept of Product Personality Assignment (PPA) (Jordan, 1997) was partially supported in the findings of Chapter 4. PPA purports that users will prefer products which match their own personalities. The agents that were chosen as the ones that were most liked matched participants in terms of neuroticism scores, but were rated more favourably in terms of extraversion scores, with the most liked agents rated as being more extraverted compared to the participant's own scores. However, the agents that were least liked were rated as being less extraverted and more neurotic when compared to the participants and the agents that were liked the most.

10.2.2 Chapter 5

In Chapter 4, participants were asked to imagine that the agents they were rating were to be used on a financial institution website. However, there was no confirmatory evidence that participants imagined the agents being used in this context, therefore it was decided to take out the factor of implied context in order to determine if there was still a general positive regard for female agents, and a negative regard for male agents. Phase 1 of this study involved participants rating a set of agents on the attributes of attractiveness and likeability. The results showed that female agents were rated as being more attractive when compared to cartoon agents, which, in turn, were rated as being more attractive compared to male agents. Although cartoon agents were liked more than female agents, and male agents were liked the least, there was a significant correlation between ratings of attractiveness and likeability for all three agent categories.

Phase 2 of this study was similar to Phase 1, but included additional male agents as the male agents in Phase 1 were all, in general, considered to be unattractive. In addition, all agents were rated on nine attributes in order to determine if the same general positive perceptions for female agents remained consistent. Results showed that female agents were considered to be significantly more attractive when compared to both cartoon agents and male agents. Ratings of attractiveness for male agents and cartoon agents were similar. In addition, female agents were rated highest on seven out of the nine attributes. The only attributes on which they were not perceived highest were friendliness and memorability; cartoon agents scored highest on both of these. The results of this study allowed a collection of agents to be chosen, which represented differing levels of attractiveness. Thus, five attractive agents, five agents of average attractiveness, and five unattractive agents were chosen. Within these five agents, each group consisted of two male agents, two female agents, and one cartoon agent.

Phase 3 of this study was conducted in order to choose a set of websites which 'matched' with the agent categories in Phase 2. That is, from a set of website, it was deemed necessary to select a set of fifteen websites, categorised in such a way that they matched the attractiveness categories of the agents. This would then allow the interaction between agent and website aesthetics to be investigated.

A set of 34 financial institutions' (e.g. bank, building society, etc) websites were chosen at random, and screen shots of the homepage used at stimuli. Each website was rated on nineteen attributes. The results showed that one bank (Barclays) was rated highest on eleven out of the nineteen attributes. However, multiple regression analysis revealed that there were no significant predictors of the overall impression of a website. Factor analysis revealed that there were three factors that explained the variance in the data; layout, appeal and aesthetics, all of which accounted for a similar amount of variance. Therefore, as there was no clear determination of the factor which best determined the overall impression of a website, and there was no a priori reason for selecting websites that scored highest on a particular attribute, the websites were chosen based on their scores across all attributes. That is, the websites that scored consistently high, average, and low across all attributes were chosen.

Thus, the results from this study provided a set of agents and websites that could be used in the next stage of experimental work to determine if there was an interaction between agent attractiveness and website 'goodness'.

10.2.3 Chapter 6

The aim of this chapter was to determine if context had an influence on perceptions of agents, and if so, the extent of this influence. In addition, agent attractiveness and gender were investigated to determine if there was an influence of these factors on perceptions of agent attributes, and also on the perceptions of 'appropriateness' of an agent for use on a financial institution website.

Each of the fifteen agents was matched with one of the fifteen websites, in three separate groups. The first group consisted of agent/website congruent pairings (i.e. attractive agents were matched with 'good' websites, average agents were matched with average websites, and unattractive agents were matched with 'bad' websites). The second and third groups consisted of agent/website incongruent pairings, so that each agent attractiveness group and website group were matched. This meant that each agent was then shown in the context of a 'good', average, and 'bad' website. Each agent was then rated on a set of nine attributes, and results showed that the influence of differing levels of website 'goodness' had no effect on ratings of agent attributes. Therefore, the three groups were collated and analysed under the factor 'context' for the remainder of the analysis of the experiment.

In addition to the factor 'context', two other conditions were included in this variable: an 'imagined' context condition, where participants were asked to imagine the agent shown being used by a financial institution on their website; and a no context condition, where participants were simply shown an image of an agent, without any explanation as how/where it would be used.

The results of this experiment revealed that, irrespective of agent attractiveness or gender, showing an agent embedded in context had a detrimental effect on ratings of agent attributes. When compared to the imagined context and no context conditions,

showing an agent in context decreased ratings on seven out of the nine attributes, and five out of the nine attributes, respectively. Ratings of agents in the imagined and no context conditions were very similar.

In order to determine the influence of agent gender, analysis was carried out, firstly irrespective of context, and then to determine the effect of context for male, female and cartoon agents. Results showed that, regardless of the context condition, there was a general positive regard for female agents. Female agents were rated higher on five out of the nine attributes when compared to male agents and cartoon agents. However, cartoon agents were considered to be more friendly and memorable when compared to both male and female agents, and more likeably and appealing when compared to male agents. Male agents were considered to be more sensible and intelligent when compared to cartoon agents. Context was also found to have an influence on ratings of female agent, with male agents being less influenced by context and cartoon agents being least influenced by context. Female agents in the imagined context condition were perceived most favourably, with ratings in the 'context' and no context conditions having a detrimental effect on ratings of seven out of the nine attributes (when compared to the imagined context condition).

Similar results were found in relation to the attractiveness of the agents. Regardless of the context condition, attractive agents were rated most favourably on eight of the nine attributes when compared to agents of average attractiveness and unattractive agents. In addition, agents of average attractiveness were rated higher on seven out of the nine attributes when compared to unattractive agents. Unattractive agents, however, were considered to be more memorable than either attractive agents or agents of average attractiveness. Context was also found to have an influence on ratings of attractive agents, and agents of average attractiveness, with unattractive agents being least influence by context. Attractive agents in the imagined context condition were perceived most favourably, with ratings in the 'context' and no context conditions having a detrimental effect on ratings of seven out of the nine attributes, and five out of the nine attribute, respectively. Similar results were found for agents of average attractiveness, with those in the imagined context condition being perceived more favourably on seven out of the nine attributes when compared to the context condition. Overall, these results suggest that embedding an agent in

context has a detrimental effect on perceptions of that agent, particularly for female agents and attractive agents.

The context (or lack of) that an agent is shown in also has an effect on the best predictors of how much an agent is liked. Multiple regression analysis showed that, overall (regardless of context condition), the best predictors of the likeability of an agent were appeal and beauty. However, in the context condition, the best predictors were trustworthiness and attractiveness; in the imagined context condition, the best predictors were appeal and attractiveness; and in the no context condition, the best predictors were friendliness and appeal. These results suggest that when an agent is not embedded in a context, then aesthetics judgements such as attractiveness and appeal determine how much an agent is liked. However, when an agent is shown in context, judgements of how much an agent can be trusted are just as important as aesthetic judgements, which suggests that users perhaps consider agents shown without context in a more superficial manner compared to when they are shown in context.

However, participants were also asked to rate each agent on how 'appropriate' they considered it for use on a financial institution website, in the context and imagined context conditions. Analysis revealed that there were no differences between the agents considered appropriate between the context and imagined context conditions. Therefore, even although context appeared to have a detrimental effect on perceptions of agents, there was still agreement about which are the most appropriate agents to use on a financial institution website. Overall, female agents were considered to be most appropriate, followed by male agents, with cartoon agents considered to be least appropriate. Consequently, even although cartoon agents were, in general, rated more highly when compared to male agents, when it comes to determining their appropriateness, cartoon agents are considered unsuitable for use on this type of website. Two out of the fifteen agents dominated appropriateness judgements – Motorola (89%) and Tmmy (81%) – both of whom were attractive, female agents. However, further analysis revealed that there were significant differences on six out of the nine attributes between Motorola and Tmmy. Motorola was considered to be significantly more intelligent, sensible, and useful when compared to Tmmy, but

Tmmy was considered to be significantly more appealing, attractive, and memorable when compared to Motorola.

Therefore, the results of this experiment suggest a number of factors that could be driving what determines the appropriateness of an agent to a financial institution website. Judgements about the 'usability' of an agent could be important, as Motorola was rated highly on the attributes of intelligence, sense, and usefulness. However, it could be 'aesthetic' factors, such as appeal and attractiveness that drives appropriateness judgements, as Tmmy was rated highly on these attributes. Finally, it could be gender stereotyping that determines the most appropriate agents. It may be the case that female agents were considered most appropriate for use on a financial institution website as bank tellers are perhaps perceived to be stereotypically female, and users may expect (and prefer) the same gender stereotyping to extend to agents.

10.2.4 Chapter 7

In order to investigate the effects of stereotyping, in particular occupational and gender stereotyping, a series of studies were carried out. Study 1 consisted of a set of fifteen occupations (five male, five female, and five neutral, based on 1991 census data) being rated on a series of questions. Results showed that out of the fifteen occupations, eight were rated as being male dominated, and four as being female dominated. In addition, the male dominated occupations were rated significantly higher on eleven out of the thirteen attributes measured when compared to the female dominated occupations. The only attributes on which the female dominated occupations were rated higher were attractiveness and preferred attractiveness. These results suggest that aesthetic factors may be more important in female dominated occupations compared to male dominated occupations. Multi-dimensional Scaling analysis was carried out to show how each occupation compared to every other occupation, on each attribute. Results showed that there was one dimension that was optimal for explaining the relationship among the occupations. Occupations were differentiated by a manual-professional dimension. Therefore, multiple regression analysis was carried out to determine the best predictors of the prestige of an occupation, and it was found that the predictors 'well-paid' and 'intelligence'

determined how prestigious an occupation was. Although the results in this study go some way in demonstrating that occupational stereotypes do exist, it was decided that the sample of occupations used was too limited, and a follow-up study including 98 occupations was carried out. The 98 occupations were first rated for gender dominance, and a more stringent criterion was imposed for determining if an occupation was male or female dominated. Results showed that out of the 98 occupations, 19 were considered to be female dominated and 12 were considered to be male dominated.

Study 2 used a sample of these occupations (6 male dominated, 6 female dominated, and 6 neutral occupations), to further investigate occupational stereotypes. A job images questionnaire (adapted from Glick et al., 1995) was used to determine the extent to which occupational stereotypes existed and if the stereotypes were based solely on gender domination. Factor analysis of the questionnaire revealed a five-factor solution, but the first two factors accounted for the majority (67%) of the variance. The first factor was interpreted as an 'emotionality' factor, with attributes such as sympathy, compassion, and affection loading highly on this factor. The second factor was interpreted as 'gender', with attributes such as communication skills, sociability, and verbal skills loading highly, and strength, spatial skills, and male physical traits loading negatively, on this factor. Consequently, 'gender' cannot be viewed simply as a 'gender dominated' factor, as it relates to the traits (both characteristic and physical) associated with males and females. When the 18 occupations were plotted against these two dimensions, it showed that the majority (12 out of 18) fell into the feminine part of the plot, including two out of the six 'male' occupations.

10.2.5 Chapter 8

The distribution of occupations along the two dimensions found in the previous study was used as a basis for the next set of experimental work. This experiment aimed to 'match-up' agents and occupations in an effort to establish the 'best' agents for different types of occupations, and to determine if there was still a general positive

view of female agents, and whether this was influenced by the occupation and/or where the occupation lay on the factor plot from the previous study.

Eight occupations were used in this experiment – two from each section of the factor plot (i.e. 2 occupations from the high emotion/male, low emotion/male, high emotion/female, and low emotion/female sections). In addition, twelve agents were selected from the attractiveness categories determined in Chapter 6: four attractive agents, four unattractive agents, and four agents of average attractiveness. Each of these attractiveness categories included two male agents and two female agents. It was decided to remove cartoon agents from this experiment as they were considered to be least appropriate for use on a financial institution website, and also because it would be difficult to determine, for cartoon agent, the extent to which gender stereotypes were driving occupational stereotypes.

Each agent was rated on a set of seventeen attributes, based purely on their physical appearance (i.e. there was no mention of the context in which the agents would be used). In addition, each of the eight occupations was rated on the same set of attributes with respect to how important each attribute was perceived for a given occupation. Participants were then shown the twelve agents and the list of eight occupations and asked to 'match-up' each occupation with one of the agents, with regards to the most appropriate, least appropriate, most preferred and least preferred.

The results of the appropriateness and preference data revealed that there was a significant positive relationship between appropriateness and preference scores for each occupation. In addition, this relationship was consistent for attractive agents, unattractive agents, agents of average attractiveness, male agents, and female agents. This suggests that an agent considered most appropriate for any given occupation is the same agent who is most preferred for that occupation.

Consideration of individual occupations revealed that female agents were regarded as most appropriate (and therefore most preferred) in the roles of cabin crew, receptionist, and shop assistant, whereas male agents were regarded as most appropriate in the roles of car dealer, labourer, and minister. There were no differences found in appropriate and preference scores between male and female

agents for the occupations of funeral director and lifeguard. These appropriateness judgements correspond, on the whole, with the division of occupations along the 'emotionality' and 'gender' dimensions described in Chapter 7. Funeral director and lifeguard both fell into the high emotion/male section of the plot, suggesting that these occupations would be suitable for either males or females (as 'high emotion' may be more associated with occupations requiring more feminine personality traits, such as compassion and affection). The main exception to the appropriateness of occupations was Minister, which was regarded as being most appropriate for a male agent, but which fell into the high emotion/feminine section of the plot. This suggests that it may be gender stereotypes which drive appropriateness judgements, and traits associate with, or considered appropriate for, the role playing being less important. This is further supported with consideration of the relationship between the perception of agent attributes and the perceived importance of occupational attributes. Although there were a few significant correlations between the attributes of the agent considered most appropriate for an occupation and the attributes of that occupation, these correlations were very weak, strengthening the idea that it may not be perceptions of the perceived traits that drive perceptions of appropriateness.

In addition, attractiveness was found to have no effect on appropriateness or preference scores (with the exception of lifeguards being considered more appropriate if they were attractive), which reinforces the idea that occupational stereotypes are driven mainly by gender, and are not based on the physical appearance of an agent. This suggests that the aesthetics of an agent plays no part in determining how appropriate an agent is for an occupation, and, perhaps more importantly, how much an agent is preferred for an occupation. It seems that the most important factor in determining agent suitability to an occupation is gender-congruence with the role.

Factor analysis revealed that the seventeen attributes were clustered differently for agents and occupations. With respect to agents, four factors were found (work-related, empathy, personal characteristics, and male gender traits), whereas six factors were found with respect to the importance of the attributes for the occupations (empathy, conscientiousness, methodical, communication, physical appearance, and intelligence). These results suggest that agent traits are viewed differently from the traits perceived to be important for an occupation, again reinforcing the idea that the

there does not need to be a 'match' in terms of perceived attributes for an agent to be suitable for a given occupation.

Consideration of individual agent attributes (irrespective of occupation) revealed that there was still a general positive regard for female agents. Female agents were rated higher on fourteen out of the seventeen attributes when compared to male agents. The only attributes on which male agents were rated higher were independence and strength. In addition, there was also still a positive general view of attractive agents, with attractive agents being rated higher than agents of average attractiveness and unattractive agents on twelve out of the seventeen attribute, and eleven out of the seventeen attributes, respectively.

However, further consideration of these results revealed that only three female agents accounted for thirteen out of the fourteen highest scores, two of which were attractive female agents, and one was an unattractive female agent. In addition, one unattractive male agent accounted for seven out of the seventeen lowest scores, and one unattractive female accounted for three of the lowest scores. However, this unattractive female agent was the same agent who accounted for four of the highest scores. Thus, there appears to be some inconsistencies, and therefore it may be possible that the high scores for a few female agents are the cause of the apparent positive general regard for female agents. Similarly, it could be a few male agents that are the cause of the generally negative perception of male agents.

However, the main conclusion from this experiment is that gender does seem to be the driving force in perceptions of agent appropriateness, and preference, for a variety of different occupations. Therefore, in order to determine if aesthetics (or agent attractiveness) does play a part in perceptions of agents, it was decided to remove gender as a factor in the final experimental study.

10.2.6 Chapter 9

As mentioned previously, gender was removed as a factor in this experiment in order to determine the role of aesthetics in the perceptions of agents. In addition, the effect

of interaction was examined in this study to establish whether perceptions of agents change after interaction.

A preliminarily study was carried out, which involved a set of 36 female agents being rated for attractiveness and age. From these agents, two were chosen: one attractive agent, and one unattractive agent, both of whom were considered to be of a similar age. These agents were then embedded in a static image of the British Airways website in order to portray the role of cabin crew (as this was found to be a 'female' occupation). In addition, a control condition was used, where no image of an agent was present. In all three conditions, a text dialogue system, controlled by a Wizard of Oz set-up, was used to generate the interaction. Pre-interaction perceptions of agents were measured on a set of twelve attributes (including eight of the attributes rated in Chapter 6, and one from each of the four factor revealed by the factor analysis in Chapter 8), for the attractive and unattractive agents. Participants in all three 'attractiveness' conditions were then given the same set of instructions for the interaction task. The task required participants to find out, via interaction with the agent, times and prices of a flight to New York, and were informed that they were free to ask any other questions or request any other information as they were interaction with a member of the cabin crew.

After interacting with the agent for 10 minutes, participants were then instructed to complete the System Usability Scale (SUS), Pleasure questionnaire, communication experience questionnaire, and agent believability questionnaire. In addition, those participants in the attractive or unattractive agent conditions were also asked to rate the agents on the same set of twelve attributes as in the pre-interaction questionnaire.

Results showed that there were few differences in perceptions of attractive and unattractive agents, both pre- and post-interaction. The only attributes on which the attractive agent scored higher than the unattractive agent were pre-interaction ratings of appeal, and post-interaction ratings of attractiveness. In addition, there were few differences between the pre- and post-interaction ratings for both the attractive and unattractive agent. For the attractive agent, post-interaction ratings of trustworthiness and usefulness increased when compared to the pre-interaction ratings. For the

unattractive agent, post-interaction ratings of appeal and usefulness increased when compared to the pre-interaction ratings.

Scores on the SUS and pleasure questionnaire revealed that there was no effect of attractiveness group on the scores. Similarly, there was no effect of attractiveness group on communication experience (which was sub-divided into four categories: face to face, involvement, co-presence, and partner evaluation). However, although the results were not significant, it is interesting to note that the agent in the control condition (i.e. when there was no image of an agent present) scored the highest mean score on the SUS, pleasure questionnaire, and all four sub-categories of the communication experience questionnaire.

In addition to the objective measures taken, subjective measures in the form of the interaction dialogue were also taken. Conversational analysis revealed that there was no effect of attractiveness group on the mean number of words used during the interaction, the mean percentage of questions asked, or on the mean percentage of statements made.

The agent believability questionnaire was used to determine those participants who thought they were interacting with a computer system, and those who thought they were interacting with another person. The data from participants who thought they were interacting with another person were removed, in order to investigate whether these data were affecting the overall results. However, analysis revealed that there was still no significant effect of attractiveness group on the SUS and pleasure questionnaire. In addition, there were no significant differences on scores on the SUS and pleasure questionnaire between those participants who believed they were interacting with a computer system and those who believed they were interacting with another person.

The findings of this experiment suggest that the attractiveness, or aesthetics, of an agent may have no impact on perceptions of that agent when interaction takes place. In fact, perceptions of system usability, pleasure and communication increased when there was no image of an agent present. This suggests that over-riding factors such as

the quality of interaction may be more salient in perceptions of agents than aesthetic judgements.

The main criticism with this experiment was that the aesthetics of the website were not measured, and this may have influenced perceptions of the quality of the system, overall. However, the findings reported in Chapter 6 suggest that the aesthetics of a website, or the overall quality of the website, do not influence perceptions of agents embedded in a website. Another influencing factor could have been the type of 'website' used, as it was not, in fact, a website, but a static image of the British Airways homepage. This may have resulted in elevated ratings on measures such as the System Usability Scale, as participants did not have to 'use' the website, but were only asked to enter text into a dialogue box.

10.3 Conclusion of experimental findings

Previous research has suggested that the way in which people interact with computers is similar to how they interact with other people (Reeves & Nass, 1996). Thus, this finding was examined in this thesis with respect to interface agents, with theories drawn from the social psychology literature. It was found that, when presented with a set of agents, participants categorised them into three distinct groups: male, female and cartoon. In general, there was a positive general perception of female agents, and a general negative perception of male agents. However, these perceptions were based on the agent being used in a context that was only implied in the study. Nevertheless, when the factor of implied context was taken out, there was still a positive regard for female agents.

Following this, a number of studies were conducted to determine participant perceptions of these different groups of agents, and their appropriateness to different contexts. Research in the areas of social psychology and marketing has suggested that there should be a 'match' between context and 'endorser' in order to optimise perceptions of the endorser. Thus, differing levels of a similar context (financial institution) were used to embed agents in (the agents differed in terms of attractiveness). However, the results obtained clearly contradict previous findings (e.g. Lawrence & Leather, 1999; Phau & Lum, 2000; Till & Busler, 1998) as the

influence of the overall impression of the website did not affect perceptions of the agents embedded in it. Furthermore, embedding an agent in context was found to have a detrimental effect on perceptions of agent attributes when compared to imagining the same agent being used in context, and even when no context is implied. Nonetheless, there was still an overall positive view of female agents, in addition to a general positive perception of attractive agents. However, it was unclear whether these consistent positive perceptions of female agents were driven mainly by perceptions of appropriateness, or whether female agents were simply liked more than male or cartoon agents, as female agents were found to be most appropriate for the role suggested by the context.

Furthermore, since the two most appropriate agents for use on a financial institution website differed in terms of aesthetics and usability factors, it was uncertain as to whether it was gender/occupational stereotypes or underlying perceptions of agent attributes that were driving the appropriateness judgements. From the studies carried out it was reported that occupational stereotyping exists along two main dimensions: emotionality and gender-type. Therefore, the influence of occupational stereotypes, and whether they extended to agents was considered. It was found that, with respect to the most appropriate agent for a given occupation, gender-congruence was the most salient aspect. However, there was still, overall, a general positive perception of female agents, and attractive agents. Consequently, in order to fully examine the effect of agent aesthetics on user perceptions, gender of the agent was removed as a factor while interaction with the agent was included as a factor since the main role of an agent is to represent something with which users can interact. Under these circumstances, it was found that there was no effect of agent attractiveness either on perceptions of agent or of system attributes. Thus, from the studies carried out the evidence strongly suggests that the aesthetics of an agent matter only if and when interaction does not take place. Arguably, then, as interaction is an essential element in the role of an agent, the most important factor in agent perception may have little to do with aesthetics or gender but may in fact be the quality of the interaction.

10.4 Overall Conclusions

Although there have been relatively few studies carried out in human-computer interaction investigating the effect of aesthetics on agents, there has been a general shift in HCI away from traditional usability aspects of design, towards pleasure, or emotion, based design. This work was interested in examining the role of these "pleasurable" aspects, such as aesthetics, to determine if the use of agents goes beyond usability.

Aesthetics have been found to influence perceived usability in ATM systems. Kurosu and Kashimura (1995) found that aesthetics were an important factor, even when users were evaluating the functional aspects of a system, while Tractinsky, Katz and Ikar (2000) found that post-use perceptions of usability were not affected by actual usability, but were affected by aesthetics, suggesting the "what is beautiful is usable" idea. Similar results were found with respect to websites, with users experiencing and judging websites, and basing their overall impression on the beauty of the website (Schenkman & Jonsson, 2000). This research supports the findings in the early chapters of this work, which suggested that the aesthetics of an agent contribute to the overall impression of that agent. Although, then, aesthetics are an important consideration in the design of a website, Tarasewich, Daniel and Griffin (2001) propose that the usability and context of the website are more important aspect, while acknowledging that aesthetics increase the 'enjoyment' factor in website use. In addition, aesthetics have been found to contribute to the overall experience when interacting with a product (Alben, 1996; Forlizzi et al., 2001), but Forlizzi suggests that there has to be a balance between functionality, usability and aesthetics to promote interaction and enhance the relationship between the user and the product. This may be one explanation as to why there was a general shift away from the importance of aesthetics in the final study of this work (Chapter 9) compared to earlier findings which suggested that there was a general preference for attractive agents. When interacting with an agent, and thus building a 'relationship', there are other aspects such as usability and functionality that are equally as important as aesthetics.

Furthermore, pleasure in design is concerned with the emotion and enjoyment experienced when using a product, and Jordan (1998; 2000) states that one of the key aspects central to pleasure-based design, and therefore the factor driving pleasure from the use of products, is not from the product itself, but from interaction and the relationship formed between the user and the product. Similar research has found that although product functions should be aesthetically pleasing, interacting with the system should contribute to the overall pleasurability of the system (Overbeeke, Djajadiningrat, Hummels & Wensven, 2000), and software design must have something, in addition to usability, to make the system novel and surprising, in order to increase the general appeal and enjoyment of use (Hassenzahl, Platz, Burmester & Lehner, 2000). It may then be the case that aesthetics are enough to invoke immediate pleasure from static images, but that interaction may educe longer-term feelings of pleasure and emotion, which Norman (2004) argues are a critical part of product design.

The relationship between the user and product/interface, or "Relationship technologies" (Lynch, Emmott & Johnson, 1999; DeAngeli, Lynch & Johnson, 2002) had been explored in relation to agents. Clarke, Jordan and Cockton (1995) found that when comparing an agent based system to a text based system, there were no differences in the ratings of usability of the systems, but the agent based system was unanimously preferred by users. However, DeAngeli et al. (2002) suggest that agents need to be socially responsive, empathic, vibrant and have 'personality' in order to maintain the user/agent relationship. Although their "Involvement Framework" suggests that aesthetics are related to pleasure, and that the physical appearance of an agent should be attractive and appealing, they also stress the importance of factors such as the ease of communication, and the amount and quality of interaction.

Catrambone et al. (2002) suggested that an increase in interaction with an agent would increase the likelihood of users thinking of it in human terms. They also proposed four main important attributes that influence agent perception: the type of agent (animated v static); amount of time on screen; role of the agent; and whether the agent is proactive or reactive. In addition, they suggest that factors such as expression, speech, gender, physical appearance, personality, competence and experience also influence user perceptions. The findings of their study suggest that the type of agent

used does not influence user perception, and that there was a general positive regard for agents. Similarly, previous experience with an agent has been found not to taint views of them, so long as the assistance received was useful and relevant (Xiao et al., 2003). In an educational setting, agents have been found to have a positive perception on the learning experience, with students being more attentive and engaged in the task (Lester et al., 1997). However, the agent used in the study carried out by Lester et al. (1997) was fully animated, which may have influence perceptions. Studies have also investigated whether the presence or absence of a face has an effect on perceptions of agents. Takeuch and Nagao (1995) found that a system was more "successful" when there was a face present, but required more effort from the user and was at times found to be distracting. Similar results were reported by Sproull et al. (1996), who found that there was no difference in the perceived intelligence of a virtual counsellor when it was represented by an animated face compared to a text only system. In addition, the text based system was rated higher on social attributes (e.g. attractiveness and friendliness) compared to the animated face. Conversely, Koda and Maes (1996) found that although a caricature male was thought of as being more intelligent when compared to a caricature dog, there was no difference in perceptions of intelligence after interaction. Furthermore, they found that there was no difference in perceptions of intelligence when there was a face present compared to when there was no face present. This supports the current research, which found that the presence of an image of an agent had no influence on perceptions of system usability, pleasure, or communication, after interaction. However, Koda and Maes also reported that, in a poker game, an opponent with a face was liked more than an invisible opponent. Similarly, participants in a study by Takeuch and Naito (1995) found a card game more entertaining when playing with a face compared to an arrow. The results of the above studies (Koda & Maes, 1996; Sproull et al., 1996), however, could be attributable to the type of agent being used (Dehn & Van Mulken, 2000); the agent in the Koda and Maes study was represented by a 2D face, whereas the agent in the Sproull et al. study was an animated 3D face. It may be the case that a more anthropomorphic agent is perceived more harshly in 'social' terms than a simple, 2D image, because it is more 'human-like'. However, although the simple, 2D face may be liked more than an 'invisible' face, interaction did not change perceptions of intelligence. This again supports the argument that the most salient aspect in the

perception of static images of agent is interaction, and how the agent looks, or if it is present at all, is irrelevant.

The findings in this thesis can also be aligned with findings in the social psychology literature. Cash and Janda (1984) state that the most influential factor in people's first impressions of others is their physical appearance, while other suggest that there is a general "beauty is good" (Dion, Berscheid & Walster, 1972) or "halo effect" (Thorndike, 1920) towards attractive people. Attractive people have also been found to be better salespeople (DeShields et al., 1996; Reingen & Kernan, 1993), and are liked more, and viewed more favourably when promoting products (Joseph, 1982), when compared to unattractive people. These theories may account for the general positive regard towards attractive agents and female agents (who were also deemed to be the most attractive). The studies carried out in Chapters 4, 5, 6, and 8 all showed images of agents, about which participants had to make judgements on a given set of attributes. In general, the results of these studies highlighted the importance of first impressions, and supported the "beauty is good" hypothesis (Dion et al., 1972). Grant et al. (2002) go on to suggest that people actually combine stereotypical characteristics of others, and then base their judgements on what that person is really like on the most salient characteristic depending on the context. This could perhaps elucidate the results found in Chapters 6 and 8, where the most salient aspect of an agent when determining the perceived appropriateness (and also preference) for a given occupation, was gender. However, the results in Chapter 6 suggested that the two most appropriate agents for use on a financial/bank website were female, although it was unclear as to whether aesthetic or usability attributes were the most influential factor driving these judgements. However, as Rao and Monroe (1989) assert, physical appearance only impacts on judgements of first impressions, and people use attributes such as appearance until they have experience with the product. It may be the case that interaction with an agent counts as 'experience', and the overall impression of the agent is based upon the interaction, which may explain why there no effect of agent attractiveness (or presence) was found in the final experiment. In the previous studies, judgements were made on images of agent, which participants only viewed on paper or as a static image on a computer screen. Although the image of the agent remained static in the last experiment, the interaction with the agent may

have been the most salient feature, and thus judgements on the physical appearance of the agent were extraneous.

10.5 Limitation of the work

The main limitation throughout each experiment was the difficulty in recruiting sufficient numbers of male participants. There were several reasons why this caused difficulties.

Firstly, the number of available males was limited as the studies were carried out predominantly among psychology students of Abertay University (which is a female dominated area) therefore it was difficult to recruit as many male participants for these studies as would have been preferred. In addition, as the studies involved agents, it was decided to exclude computing students (which is a male dominated area) from the sample in order to reduce any prior knowledge of agents affecting results.

Additionally, a high percentage of participants failed to turn up to take part in the experiments (particularly the experiment in Chapter 9) and therefore the number of recruited participants was at least 20% lower than the target sample number. Although there was a considerably high drop out or "no-show" rate of male participants it was much easier to recruit female participants to take part in the experiments when this occurred. Although a target number of male participants was set at the beginning of the study it was not always possible to achieve this number (due to the reasons outlined above) and therefore it was difficult keep the proportion of male and female participants equal. Although some participant gender differences were found, which may have been attributable to the uneven sample sizes, there was a consistent, significant, positive correlation between male and female participant scores. In addition, the majority of the social psychology literature on rating others, and in particular perceptions of the physical attractiveness stereotypes, suggests that males and females rate others similarly (Eagly et al., 1991).

Another problem with the study described in Chapters 4 and 5 was that there appeared to be more 'attractive' female agents compared to male agents available. Throughout

this work, data suggested that there was a general positive regard for female agents. Although this finding may be accurate, it is difficult to make a strong claim to that effect as it may simply have been a lack of 'attractive' male agents used in the sample. Additionally, previous research has suggested that there is a general agreed standard for female attractiveness, but not for male attractiveness, which again may have influenced the high perceptions of female agents.

10.6 Reflections on lessons learned

The most important lesson learned during the undertaking of this thesis, and perhaps most evidently so from Chapter 9, was the consequence of having a limited sample size. Although the findings of Chapter 9 were non-significant there was very low statistical power, thus the conclusions drawn may have been as a result of no true effect or because of the small sample size (38 participants). In addition, the use of standardised questionnaires, particularly with regard to rating agents, may have been beneficial to ensure the experiments carried out were more ecologically valid.

10.7 Future research

In order to follow up, investigate further, and extend many of the findings that have been reported during the course of this thesis a number of future studies are proposed here.

The final study (interaction study in Chapter 9) examined issues of agent aesthetics along with agent interaction and produced the most interesting and important result. It appeared that agent aesthetics did not significantly affect perceptions of agents, or perceptions of the system. In fact, on average, in the condition where no image of an agent was present the system was perceived to be easier and more pleasurable to use, compared to when an image of an agent was present. The interactions were also examined and it was found that there was no effect of agent aesthetics on perceived quality. Importantly, when the interactions were investigated, there were no differences between agent present and agent absent conditions. The findings from this study would, therefore suggest that the quality of the interaction may be more important than the physical appearance of the agent (both in terms of being present or

absent and of aesthetics when present), and so long as there is gender-congruence between the agent and the role it is portraying. However, the results from previous chapters (4, 5, 6, and 8) suggested that attractive agents were preferred over agents of average attractiveness and unattractive agents.

This finding needs to be investigated further, and future studies will examine whether the perceived quality of interaction remains consistent for a gender-incongruent agent. If a male agent had been used in the final study, the effect of aesthetics might have played a more salient role in the perception of the agent. As Cann (1993) found, gender-inconsistencies are more likely to affect perceptions of incompetence, and if the quality of interaction was poor then attractiveness could matter (Landy & Sigall, 1974). It could be that a male agent in a typically 'female' role would have to be attractive in order to compensate for the potential perceptions of poor performance.

Additionally, future studies will also examine the effect of social interaction on the perception of agents. Although the final study was concerned with interaction, it was task-orientated, with participants being given specific instructions on the information they had to find out. On the basis of the social psychological literature, it could be hypothesised that if participants were interacting with an agent in a social manner there would be an effect of agent aesthetics. For instance, Dion, Berscheid and Walster (1972) reported that attractive people are considered to be more socially desirable compared to unattractive people, and a number of researchers have found that the 'beauty is good' stereotype is strongest on judgements of social competence (Bassili, 1981; Dion, 1981, 1986; Eagly, Ashmore, Makhijani & Longo, 1991). In addition, it might be the case that perceptions of agents change over time and therefore, if there is an effect of aesthetics, it may only be on the first impressions of the agent.

Finally, this work was concerned only with static images of agents. There are many ways in which an agent can be represented, ranging from a simple, static image, to a fully embodied, 3D, talking agent, which, through the use of inflection, gaze and facial expression can convey emotions and affect. The effect of aesthetics needs to be further examined in these more 'life-like' agents to determine if what is beautiful is really good.

References

Abramowitz, L. A., & O'Grady, K. E. (1991). The impact of gender, physical attractiveness, and intelligence on the perception of peer counselors. *Journal of Psychology, 125,* 311-326.

Adams, G. R. (1977). Physical Attractiveness Research. Toward a Developmental Social Psychology of Beauty. *Human Development, 20,* 217-239.

Adams, G. R., & LaVoie, J. C. (1975). Parental expectations of educational and personal-social performance and childrearing patterns as a function of attractiveness, sex, and conduct of the child. *Child Study Journal, 5,* 125-142.

Ahearne, M., Gruen, T. W., & Burke C.J. (1999). If looks could sell: Moderation and mediation of the attractiveness effect on salesperson performance. *International Journal of Research in Marketing, 16,* 269-284.

Alben, L. (1996). Quality of Experience: Defining the Criteria for Effective Interaction Design. *Interations, 3* (3), 11-15.

Algoe, S. B., Buswell, B. N., & DeLamater, J. D. (2000). Gender and Job Status as Contextual Cues for the Interpretation of Facial Expression of Emotion. *Sex Roles, 42,* 183 – 208.

Asch, S. E. (1946). Forming Impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258-290.

Baker, M. & Churchill, G. (1977). The Impact of Physically Attractive Models on Advertising Evaluation. *Journal of Marketing Research,* 14, 538-555.

Baron, R. A., & Byrne, D. (1994). *Social Psychology. Understanding Human Interaction* (7th ed.). Massachusetts: Allyn and Bacon.

Bassili, J. N. (1981). The Attractiveness Stereotype: Goodness or Glamour? *Basic and Applied Social Psychology, 2*, 235-252.

Belch, G. E., & Belch, M. A. (1998). *Advertising and promotion: An integrated marketing communications perspective.* (4th ed.). Boston: Irwin McGraw-Hill.

Bem, S. L. (1974). The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 42, 155-162.

Berscheid, E. (1985). Interpersonal Attraction. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*. New York: Random House.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, I. L. (Eds.), *Usability evaluation in industry*, 189-194. London, UK: *Taylor & Francis.*

Buisine, S., & Martin, J. C. (2003). *Experimental evaluation of bi-directional multimodal interaction with conversational agents.* Paper presented at the Proceedings of INTERACT '2003, Zürich, Switzerland.

Byrne, D. (1961). Interpersonal Attraction and attitude similarity. *Journal of Abnormal and Social Psychology, 62*, 713-715.

Caballero, M. and Pride, W. (1984), Selected effects of salesperson sex and attractiveness of direct mail advertisements. *Journal of Marketing, 48*, 94-100.

Cann, A. (1993). Evaluative Expectations and the Gender Schema: Is Failed Inconsistency Better? *Sex Roles, 28*, 667-678.

Carroll, J. M., & Russell, J. A, (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70* (2), 205 - 218.

Cash, T., & Janda, L. H. (1984). The eye of the beholder. *Psychology Today, 18,* 46-52.

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM, 43*, 70-78.

Cassell, J., & Vilhjalmsson, H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. *Autonomous Agents and Multi-Agent Systems, 2* (1), 45-64.

Catrambone, R., Stasko, J., & Xiao, J. (2002). *Anthropomorphic agents as a user interface paradigm: experimental findings and a framework for research.* Proceedings of the 24th Annual Conference of the Cognitive Science Society, 166-171.

Cavazza, M., Charles, F., & Mead, S. J. (2002). *Interacting with virtual characters in interactive storytelling.* Paper presented at the AAMAS '02, Bologna, Italy.

Cejka, M. A., & Eagly, A. H. (1999). Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment. *Personality and Social Psychology Bulletin, 25* (4), 413-423.

Churchill, G. A., Ford, N, M., Hartley, S. W., & Walker, O. C. (1985). The Determinants of Salesperson Performance: A Meta-Analysis. *Journal of Marketing Research, 22* (2), 103-118.

Clarebout, G., Elen, J., Johnson, W. L., & Shaw, E. (2002). Animated pedagogical agents: An opportunity to be grasped? *Journal of Educational Multimedia and Hypermedia, 11* (3), 267-286.

Clarke, S., Jordan, P. W., & Cockton, G. (1995). Applying Aristotle's Theory of Poetics to Design. In S. Robertson (Ed.), *Contemporary Ergonomics*. London: Taylor and Francis.

Creusen, M. E. H., & Snelders, H. M. J. J. (2002). Product appearance and consumer pleasure. In WS Green & PW Jordan (Eds.), *Pleasure with products: beyond usability* (pp. 69-75). London: Taylor & Francis.

Crozier, R. (1994). *Manufactured Pleasures*. Manchester: Manchester University Press.

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz Studies - why and how. *Knowledge-Based Systems, 6* (4), 258-266.

De Angeli, A., Lynch, P., & Johnson, G. I. (2002). Pleasure versus Efficiency in User Interfaces: Towards an Involvement Framework. In W. Green & P. W. Jordan (Eds.), *Pleasure with Products: Beyond Usability*. London: Taylor and Francis.

Debevec, K., Madden, T. J., & Kernan, J. B. (1986). Physical attractiveness, message evaluation, and compliance: A structural examination. *Psychological Reports, 58*, 503-508.

DeShields, O., Kara, A., & Kaynak, E. (1996). Effects in purchase decisions: The impact of physical attractiveness and accent of salesperson. *International Journal of Research in Marketing, 13* (1), 89 – 101.

De Wulf, K., Schillewaert, N., Muylle, S., & Rangarajan, D. (2006). The role of pleasure in web site success. *Information and Management, 43*, 434 - 446

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies, 52*, 1-22.

Dion, K. K. (1973). Young Children's stereotyping of facial attractiveness. *Developmental Psychology, 9*, 183-188.

Dion, K. K. (1981). Physical Attractiveness, sex roles and heterosexual attraction. In M. Cook (Ed.), *The bases of human sexual attraction*. London: Academic Press.

Dion, K. K. (1986). Stereotyping based on physical attractiveness: Issues and conceptual perspectives. In C. P. Herman & M. P. Zanna & E. T. Higgens (Eds.), *Physical Appearance, stigma and social behavior: The Ontario Syposium*. Hillsdale, NJ: Erlbaum.

Dion, K. K., & Berscheid, E. (1974). Physical attractiveness and peer perception among children. *Sociometry, 37*, 1-12.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is Beautiful is Good. *Journal of Personality and Social Psychology, 24* (3), 285-290.

Dwyer, F.R., Schurr, P.H. and Oh, S. (1987). Developing buyer-seller relationships. *Journal of Marketing,* 51 (2), 11-27.

Eagly, A. H. (1987). *Sex Differences in Social Behavior: A Social-Role Interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is Beautiful is Good, But...: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype. *Psychological Bulletin, 110* (1), 109-128.

Erwin, P. G. (1993). First names and perceptions of physical attractiveness. *The Journal of Psychology, 127*, 625-631.

Erwin, P. G. (1999). Attractiveness of first names and academic achievement. *The Journal of Psychology, 133* (6), 617-620.

Erwin, P. G., & Calev, A. (1984). The influence of Christian name stereotypes on the marking of children's essays. *British Journal of Educational Psychology, 54*, 223-227.

Evans, J. (2007, February 7). iPods dominate Amazon's US and UK sales charts. Retrieved May 31, 2007, from http://www.macworld.co.uk/ipod-itunes/news/index.cfm?newsid=17114

Fielding, D., Fraser, M., Logan, B., & Benford, S. (2004). Extending game participation with embodied reporting agents. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology ACE '04*

Forlizzi, J. & Ford, S. (2002). The building blocks of experience: An early framework for interaction designers. *Designing Interactive Systems 2000 Conference Proceedings*, New York, NY, 419–423.

Forlizzi, J., Hirsch, T., Hyder, E., & Goetz, J. (2001). *Designing Pleasurable Technology for Elders.* Paper presented at the INCLUDE, International Conference on Inclusive Design and Communications, London, UK.

Garau, M, Slater, M., Bee, S. & Sasse, M. A. (2001). The impact of eye gaze on communication using humanoid avatars. In J. A. Jacko, A. Sears, M. Beaudouin-Lafon & R. Jacob (Eds.), *Proceedings of CHI 2001* (pp. 309-316). New York: ACM Press.

Glick, P. (1991). Trait-based and Sex-based Discrimination in Occupational Prestige, Occupational Salary, and Hiring. *Sex Roles, 25*, 351-378.

Glick, P., Wilk, K., & Perreault, M. (1995). Images of Occupations: Components of Gender and Status in Occupational Stereotypes. *Sex Roles, 32* (9/10), 565-582.

Goldberg, M. E., & Hartwick, J. (1990). The effects of advertiser reputation and extremity of advertising claim on advertising effectiveness. *Journal of Consumer Research,* 17, 172–179.

Goldman, W., & Lewis, P. (1977). Beautiful is good: Evidence that the physically attractive are more socially skilful. *Journal of Experimental Social Psychology, 13*, 125-130.

Goleman, D. (1996). *Emotional Intelligence: Why It Can Matter More than IQ.* London; Bloomsbury Publishing.

Grant, M. J., Button, C. M., Hannah, T. E., & Ross, A. S. (2002). Uncovering the Multidimensional Nature of Stereotype Inferences: A Within-Participants Study of Gender, Age, and Physical Attractiveness. *Current Research in Social Psychology, 8* (2), 19 – 38.

Gruen, D., Sinder, C., Boettner, C., & Rich, C. (1999, 15-20 May 1999). *A Collaborative Assistant for EMail.* Paper presented at the CHI '99, Pittsburgh, PA.

Hamilton, W. (2002). Personal Communication.

Harari, H., & McDavid, J. W. (1973). Name stereotypes and teacher expectations. *Journal of Educational Psychology, 65*, 222-225.

Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). *Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal.* Paper presented at the CHI 2000.

Hayes, N. (1993). *Principles of Social Psychology.* UK: Lawrence Erlbaum Associates Ltd.

Heilman, M. E., & Saruwatari, L. R. (1979). When beauty is beastly: The effects of appearance and sex on evaluations of job applicants for managerial and non-managerial jobs. *Organizational Behavior and Human Performance, 23*, 360-372.

Hook, K., Persson, P., & Sjolinder, M. (1999). *Measuring experience of interactive characters.* W. S. Green and P. W. Jordan (eds.) Pleasure with Products: Beyond Usability, Taylor and Francis, London.

Horai, J., Naccari, N. and Fatoullah, E. (1974). The Effects of Expertise and Physical Attractiveness upon Opinion Agreement and Liking. *Sociometry*, 37, 601-606.

Hudson, W. (2000). Web Evolution: Is HCI an Endangered Species? *SIGCHI Bulletin, November/December*, 9.

Hummels, C. (Ed.). (1999). *Engaging contexts to evoke experiences*. Delft, The Netherlands.

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies, 53*, 251-267.

Isen, A. (2000). *Positive Affect and Decision-making.* In M. Lewis and J. Haviland Jones (eds), *Handbook of emotions* (2nd ed. pp 417-36).

Iwata, S., Takahiro, M., & Morihara, T. (1999). Human-machine Interface Using Humaniod Cartoon Character. *FUJITSU. Sci. Tech., 35* (2), 165-173.

Jones, E. E. (1990). *Interpersonal Perception.* New York: W. H. Freeman and Company.

Jordan, P. W. (1997a). Putting the pleasure into products. *IEE Review, November 1997*, 249-252.

Jordan, P. W. (1997b). Products as personalities. In: S. A. Robertson (Ed.), *Contemporary Ergonomics 1997.* London: Taylor and Francis.

Jordan, P. W. (1998). Human Factors for pleasure in product use. *Applied Ergonomics, 29* (1), 25-33.

Jordan, P. W. (1999). Pleasure with products: Human factors for body, mind and soul. In: W. S. Green and P. W. Jordan (Eds.), *Human Factors in Product Design: Current Practice and Future Trends*, London: Taylor and Francis, 206 – 217.

Jordan, P. W. (2000). *Designing Pleasurable Products*: Taylor and Francis.

Jordan, P.W., & Servaes, M. (1995). Pleasure in product use: beyond usability. In: S. Robertson (Ed.), Contemporary Ergonomics 1995. London: Taylor and Francis.

Joseph, W. (1982). The credibility of physically attractive communicators: A review. *Journal of Advertising, 11,* 15-24.

Kahle, L. R. & Homer, P. M. (1985). Physical attractiveness of the celebrity endorser: A social adaptation perspective. *Journal of Consumer Research*, 11, 954-961.

Kelly, G. A. (1955). *The Psychology of Personal Constructs.* New York: Norton.

Kelly, K. (1999). *New rules for the new economy: 10 ways the network economy is changing everything*. London: Forth Estate.

Kerchoff, A. C., & Davis, K. I. (1962). Value consensus and need complementarity in mate selection. *American Sociological Review, 27*, 295-303.

King, W. J., & Ohya, J. (1996). *The representation of agents: anthropomorphism, agency, and intelligence.* Paper presented at the Proceedings of CHI '96 Conference Companion, Vancouver, Canada.

Koda, T. and Maes, P. (1996). Agents with Faces: The effects of personification of agents. Proceedings of Human-Computer Interaction. London, UK., 239 - 245.

Kurosu, M., & Kashimura, K. (1995). *Apparent Usability vs. Inherent Usability.* Paper presented at the CHI '95.

Landauer, T. K. (1987). Psychology as a mother of invention. *Proceedings ACM CHI+GI '87*, 333-335.

Landy, D., & Sigall, H. (1974). Beauty is talent. Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*, 299-304.

Langmeyer, L., & Shank, M. (1994). Managing Beauty - Products and People. *Journal of Product and Brand Management, 3* (3), 27-38.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115-121.

Langlois, J. H., Roggman, L. A., & Musselman, L. (1994). What is average and what is not average about attractive faces? *Psychological Science, 5,* 214-220.

Lanier, J. (1995). Agents of alienation. *Interactions, 2* (3), 66-72.

Larose, H., J. Tracy & S.J. McKelvie (1992). Effects of gender on the physical attractiveness stereotype. *Journal of Psychology*, 127 (6), 677–680.

Laurel, B. (Ed.). (1990). *Interface Agents: Metaphors with Character.* Reading, MA: Addison-Wesley.

Laurel, B. (1993). *Computers as Theatre*: Addison-Wesley Publishing Company, Inc.

Lawrence, C., & Leather, P. (1999). Stereotypical Processing: The Role of Environmental Context. *Journal of Environmental Psychology, 19*, 383-395.

Lester, J. C., Converse, S. A., Stone, B. A., Kahler, S. H., & Barlow, S. T. (1997). *Animated Pedagogical Agents and Problem-Solving Effectiveness: A Large-Scale Empirical Evaluation.* Paper presented at the Eighth World Conference on Artificial Intelligence in Education, Kobe, Japan.

Little, A. C., Burt, D. M., Penton-Voak, I. S., & Perrett, D. I. (2001). Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society of London. Biology: 268*, 39-44

Lynch, P., Emmot, S. E., &Johnson, G. I., (1999). The NCR knowledge lab. In: S. Brewster, A. Cawsey and G. Cockton (Eds.), Human-Computer Interaction: Interact '99, Proceedings. IOS Press/IFIP, 229-230.

Maier, N. R. F. (1955). *Psychology in Industry.* New York: McGraw Hill.

Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology, 52* (3), 489-499.

Maulsby, D., Greenberg, S., & Mander, R. (1993). *Prototyping an Intelligent Agent through Wizard of Oz.* Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands.

McCracken, G. (1989). Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research*, 16 , 310-21.

Mitchell, A.A. & Olson, J.C. (1981). Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attribute? *Journal of Marketing Research*, 18, 318-332.

Nass, C., Steuer, J., & Tauber, E. R. (1994). *Computers are Social Actors.* Paper presented at the CHI '94, Boston, Massachusetts, USA.

Nass, C., Isbister, K., & Lee, E. J. (2000). Truth is beauty: Researching embodied conversational agents. In: J. Cassell (Ed.), Embodied conversational agents. Cambridge, MA: MIT Press.

Nielsen, J. (1993). *Usability Engineering.* San Diego, CA: Morgan Kaufmann.

Norman, D. A. (2004). Emotional Design: Why We Love (or Hate) Everyday Things. New York: Basic Books.

Oh, W., & Khong, P. W. (2003). Competitive advantage through pleasurable products. *Proceedings of The International Conference on Designing Pleasurable Products and Interfaces, Pittsburgh, PA,* 87 – 91.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising, 19* (3), 39-52.

Overbeeke, K., Djajadiningrat, T., Hummels, C., & Wensveen, S. (Eds.). (2002). *Beauty in Usability: Forget about ease of use.* London: Taylor and Francis.

Perrett, D. I., May, K., & Yoshikawa, S. (1994). Attractive characteristics of female faces: preference for non-average shape. *Nature, Lond., 368*, 239-242.

Phau, I., & Lum, L. (2000). Effects of 'physical attractiveness' in the evaluation of print advertisements. *Asia Pacific Journal of Marketing and Logistics, 12* (4), 41-59.

Rao, A., & Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers perceptions of product quality, and integrative review. *Journal of Marketing Research, 26*, 351-357.

Reeves, B., & Nass, C. (1996). *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge: CSLI Publications/Cambridge University Press.

Reingen, P. H., & Kernan, J. B. (1993). Social Perception and Interpersonal Influence: Some Consequences of the Physical Attractiveness Stereotype in a Personal Selling Setting, *Journal of Consumer Psychology*, 2 (1), 25-38.

Rhodes, B. J. (2000). *Margin Notes: Building a contextually aware associative memory.* Paper presented at the International Conference on Intelligent User Interfaces, IUI 00, New Orleans, LA.

Rickenberg, R., & Reeves, B. (2000). The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. *CHI Letters, 2* (1).

Rosenberg, S. (1977). New approaches to the analysis of personal constructs in person perception. In A. L. Land & J. K. Cole (Eds.), *Nebraska Symposium on Motivation* (24). Lincoln: University of Nebraska Press.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A Multidimensional Approach to the Structure of Personality Impressions. *Journal of Personality and Social Psychology, 9*(4), 283-294.

Russo, B., & de Moraes, A. (2003). The lack of usability in design icons. An affective case study about Juicy Salif. *Proceedings of The International Conference on Designing Pleasurable Products and Interfaces, Pittsburgh, PA,* 146-147.

Ruttkay, Z., Dormann, C., & Noot, H. (2002). *Evaluating ECAs - What and How?* Paper presented at the Workshop on "Embodied conversational agents - let's specify and

evaluated them!" in conjunction with The First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '02), Bologna, Italy.

Salber, D., & Coutaz, J. (1993). *A Wizard of Oz Platform for the Study of Multimodal Systems.* Paper presented at the INTERCHI'93.

Schenkman, B. N., & Jonsson, F. U. (2000). Aesthetics and Preferences of Web Pages. *Behaviour and Information Technology, 19* (5), 367-377.

Scholtz, J., Laskowski, S., & Downey, L. (1998). *Developing Usability Tools and Techniques for Designing and Testing Websites.* Paper presented at the 4th Conference on Human Factors and the Web, Basking Ridge, New Jersey.

Selker, T. (1994). Coach: A teaching agent that learns. *Communications of the ACM, 37* (3), 92-99.

Shackel, B. (Ed.). (1991). *Usability: Context, framework, definition, design and evaluation.* Cambridge: Cambridge University Press.

Shahani-Denning, C., & Plumitallo, D. (1993). *The influence of physical attractiveness and gender on disciplinary decisions.* Paper presented at the Fifth Annual Convention of the American Psychological Society, Chicago, IL.

Shneiderman, B. (Ed.). (1997). *Direct manipulation versus agents: Paths to predictable, controllable, and comprehensive interfaces.* Cambridge, MA: MIT Press.

Shneiderman, B., & Maes, P. (1997). Direct manipulation vs. interface agents. *Interactions, 4*(6), 42-61.

Solomon, M. R., Ashmore, R. D., & Longo, L. C. (1992). The Beauty Match-Up Hypothesis: Congruence Between Types of Beauty and Product Images in Advertising. *Journal of Advertising, 21* (4), 23-34.

Spence, J. T. (1993). Gender-related traits and gender ideology: Evidence for a multi-factorial theory. *Journal of Personality and Social Psychology, 64,* 624 – 635.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human–Computer Interaction, 11,* 97–124.

Stolterman, E. (1994). Guidelines or aesthetics - design learning strategies. *Design studies*, 15 (4), 448-558.

Takeuchi, Y., & Nagao, K. (1995). *Situated Facial Displays: Towards Social Interaction.* Paper presented at the Proceedings of CHI '95, New York.

Tannenbaum, R. S. (1998). *Theoretical Foundations of Multimedia.* New York: Computer Science Press.

Tarasewich, P., Daniel, H. Z., & Griffin, H. E. (2001). Aesthetics and Web Site Design. *Quarterly Journal of Electronic Commerce, 2* (1), 67-81.

Thorndike, E. L. (1920). "A Constant Error on Psychological Rating." *Journal of Applied Psychology*, vol. IV, 25-29

Tiger, L. (1992). *The pursuit of pleasure.* Boston: Little, Brown & Company

Till, B. D., & Busler, M. (1998). Matching Products with Endorsers: Attractiveness versus Expertise. *Journal of Consumer Marketing, 15* (6), 576-586.

Tractinsky, N. (1997). *Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues.* Paper presented at the CHI' 97, Atlanta, GA, USA.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable [user interface design]. *Interacting-with-Computers, 13* (2), 127-145.

Walker, J. H., Sproull, L., & Subramani, R. (1994). *Using a Human Face in an Interface.* Paper presented at CHI '94, Boston, Massachusetts, USA.

Wilson, J., & Rosenberg, D. (Eds.). (1988). *Rapid Prototyping for User Interface Design.* New York: North-Holland.

Wilson, K. (2002). Evaluating images of virtual agents. Poster session presented at the Conference on Human Factors in Computing Systems (CHI 2002), Minneapolis, MN.

Xiao, J., Catrambone, R., & Stasko, J., (2003). Be Quiet? Evaluating proactive and reactive user interface assistants. Technical Report GIT-GVU-03-03, February 2003.

Xiao, J., Stasko, J., & Catrambone, R. (2002). *Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation.* Paper presented at the First International Joint Conference on Autonomous Agents & Multi-Agent Systems, Bologna, Italy.

Appendices

Appendix 1

Agents used in Chapter 4

Appendix 2

Instructions and example of Task 1, Chapter 4

Subject No:_____

Instructions for Task 1

The purpose of this experiment is to select a character to be used as the face for a financial company. It will be used as an interface agent which will interact with users on the company web-site.

You are going to be presented with 2 characters on the screen. Above each pair either 'Similarity' or 'Preference' will be written. Your task is to judge each pair of characters on how similar they are and then to state your preference between them.

The similarity or preference scale will be displayed below each pair.

If the word is 'Similarity':

- Judge the characters on how similar they are on a scale from 1 to 10, where 1 is 'Fully Dissimilar' and 10 is 'Fully Similar'

If the word is 'Preference':

- Judge the character on your personal preference on a scale from 1 to 10, where 1 is 'Left hand character preferred completely' and 10 is 'Right hand character preferred completely'.

Examples of these are shown below.

Your judgements are made by pressing the corresponding number on the keyboard.
Note: The '0' key represents 10

There is no time limit on which to make your judgements.

Fully Dissimilar	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Fully Similar
Left hand character preferred completely	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Right hand character preferred completely

Similarity

Fully Dissimilar 1 2 3 4 5 6 7 8 9 10 **Fully Similar**

Preference

Left Hand Character Preferred Completely 1 2 3 4 5 6 7 8 9 10 **Right Hand Character Preferred Completely**

Appendix 3

Instructions for Task 2, Chapter 4

Subject No: _____

Instructions for Task 2

You are going to be presented with 14 characters one at a time on the screen.
One of these characters will be used as an interactive computer agent for a financial company on their web-site.

Your task is to rate each character in turn on the questionnaires provided. Please indicate at the beginning of the questionnaire which character you are rating. When you have finished rating a character, press any key to display the next character.

Please rate each of the characters shown by circling a number on the scales below.

Character ____

Does the character appear to be:

Beautiful	1	2	3	4	5	6	7	8	9	10	Ugly
Unpleasant	1	2	3	4	5	6	7	8	9	10	Pleasant
Appealing	1	2	3	4	5	6	7	8	9	10	Unappealing
Stupid	1	2	3	4	5	6	7	8	9	10	Intelligent
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Foolish	1	2	3	4	5	6	7	8	9	10	Sensible
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Inappropriate	1	2	3	4	5	6	7	8	9	10	Appropriate
Memorable	1	2	3	4	5	6	7	8	9	10	Forgettable

What is your overall impression of the character?

Excellent	1	2	3	4	5	6	7	8	9	10	Poor

Appendix 4

EPI Questionnaire

Due to copyright reasons, the Eysenck Personality Inventory (EPI) could not be reproduced. The questionnaire is available from the publisher, Hodder & Stoughton.

Appendix 5

Mann-Whitney summary for task 2 – male and female differences

Attribute	Mann-Whitney statistical summary
Beauty	U = 95, N1 = 14, N2 = 14, p = 0.89
Pleasant	U = 94.5, N1 = 14, N2 = 14, p = 0.87
Appeal	U = 91, N1 = 14, N2 = 14, p = 0.75
Intelligence	U = 92.5, N1 = 14, N2 = 14, p = 0.80
Trust	U = 89, N1 = 14, N2 = 14, p = 0.68
Sensible	U = 92.5, N1 = 14, N2 = 14, p = 0.80
Useful	U = 95.5, N1 = 14, N2 = 14, p = 0.91
Appropriate	U = 89, N1 = 14, N2 = 14, p = 0.68
Memorable	U = 62.5, N1 = 14, N2 = 14, p = 0.10
Overall Impression	U = 90.5, N1 = 14, N2 = 14, p = 0.73

Appendix 6

Table of mean scores for all agents

Agent	Beauty	Pleasant	Appeal	Intelligence	Trust
A	6.3750	5.6875	6.8750	5.6875	6.8750
B	4.9375	4.3750	4.6875	3.9375	3.8125
C	4.0625	3.3750	4.0000	3.0625	3.1250
D	5.0625	4.6250	5.3125	4.8750	4.8125
E	5.0625	4.7500	4.6875	4.8750	4.5625
F	7.9375	6.5000	7.2500	5.8750	5.3125
G	7.6250	6.5625	6.9375	7.8750	5.9375
H	6.1250	4.1875	4.8125	6.1875	5.5625
I	5.8750	3.0625	4.5000	7.3125	5.0625
J	6.2500	5.9375	6.5000	7.4375	6.4375
K	8.1875	7.3125	7.7500	8.5000	7.4375
L	6.2500	6.1250	6.8125	3.5625	5.0625
M	2.0625	2.3750	2.1875	4.5000	4.5000
N	7.6875	7.5625	8.1250	5.3125	5.8125

Agent	Sensible	Useful	Appropriate	Memorable	Overall Impression
A	5.4375	5.5625	5.7500	6.5000	6.3125
B	3.6250	3.8125	3.3125	5.9375	4.6250
C	2.6250	2.8125	2.6250	5.0625	3.9375
D	4.8750	4.8750	4.3125	5.5000	5.9375
E	5.5000	4.7500	5.1875	3.6875	5.3750
F	6.1250	5.6875	6.2500	5.9375	7.0625
G	8.0625	7.5000	8.1875	4.8125	7.3125
H	6.7500	6.0000	7.1250	4.6875	5.7500
I	7.2500	5.9375	7.3750	3.3750	4.8125
J	8.0000	6.5625	7.5625	3.7500	6.9375
K	7.7500	8.5000	8.5625	6.5000	8.3750
L	3.6250	4.5000	4.0625	6.0625	6.1875
M	4.5625	3.5625	4.7500	2.9375	3.8125
N	5.0000	6.0625	5.8750	5.7500	7.4375

Appendix 7

T-test summary for EPI – male and female differences

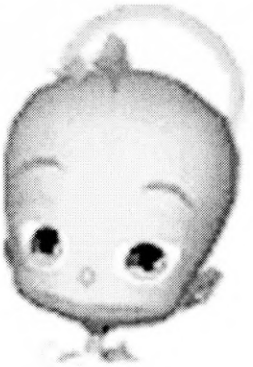
		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Self Extrovert Score	Equal variances assumed	.810	14	.432
Agent Like Extrovert Score	Equal variances assumed	.702	14	.494
Agent Hated Extrovert Score	Equal variances assumed	.081	14	.936
Self Neurotic Score	Equal variances assumed	-.113	14	.912
Agent Liked Neurotic Score	Equal variances assumed	.995	14	.337
Agent Hated Neurotic Score	Equal variances assumed	.323	14	.752

Appendix 8

Agents and rating scales used in Chapter 5, phase 1

Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

Like 1 2 3 4 5 6 7 8 9 10 Dislike



Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

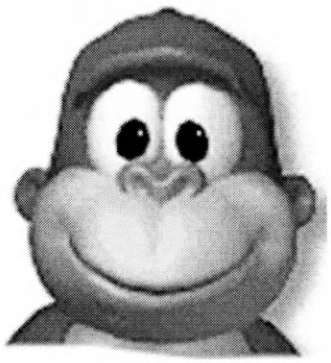
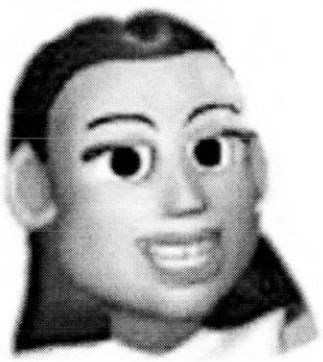
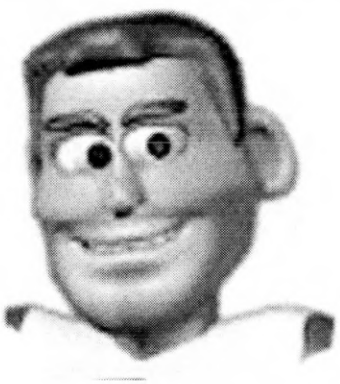
Like 1 2 3 4 5 6 7 8 9 10 Dislike

Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

Like 1 2 3 4 5 6 7 8 9 10 Dislike

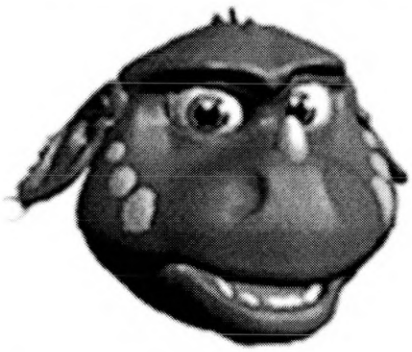
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

Like 1 2 3 4 5 6 7 8 9 10 Dislike

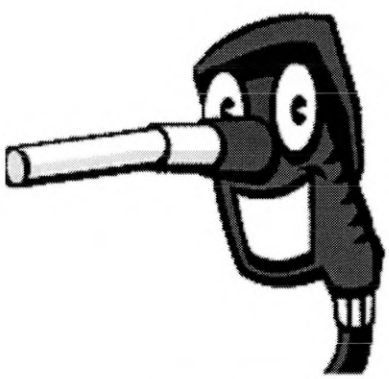
	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike
	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

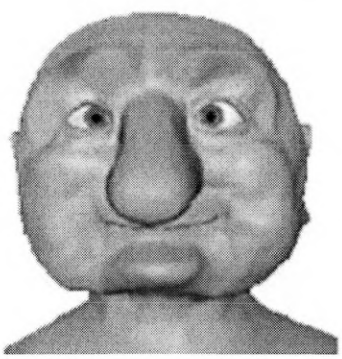
Appendix 8 continued

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

Appendix 8 continued

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike
	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike
	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

Appendix 8 continued

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike
	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

Appendix 8 continued

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike
	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

Appendix 8 continued

Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

Like 1 2 3 4 5 6 7 8 9 10 Dislike

Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

Like 1 2 3 4 5 6 7 8 9 10 Dislike

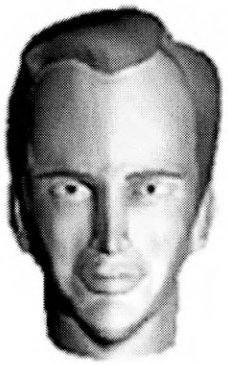
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

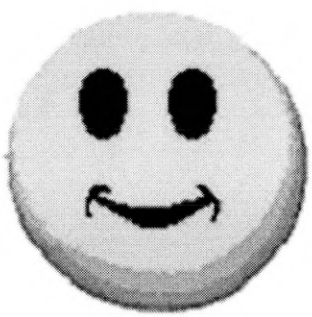
Like 1 2 3 4 5 6 7 8 9 10 Dislike

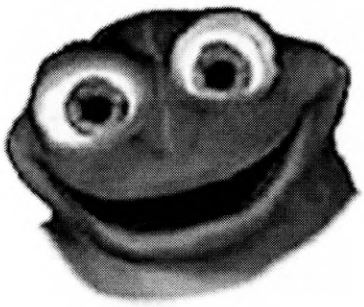
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive

Like 1 2 3 4 5 6 7 8 9 10 Dislike

Appendix 8 continued

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

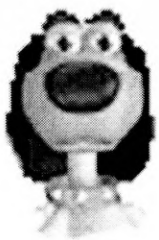
Appendix 8 continued

	
Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike	Attractive 1 2 3 4 5 6 7 8 9 10 Unattractive Like 1 2 3 4 5 6 7 8 9 10 Dislike

Appendix 9

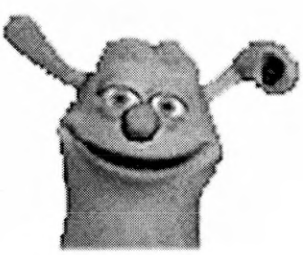
30 agents used in Chapter 5, phase 2

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike



Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike



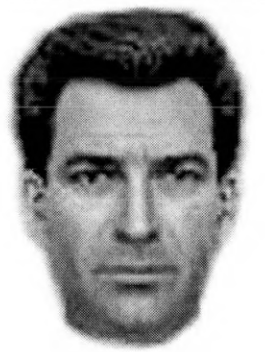
Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Appendix 9 continued

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

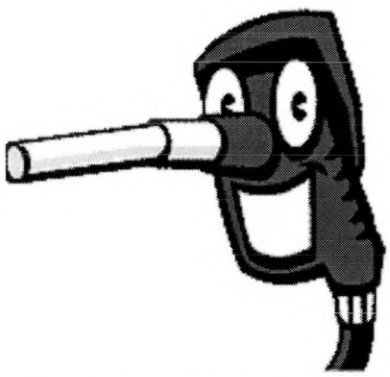


Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Appendix 9 continued

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

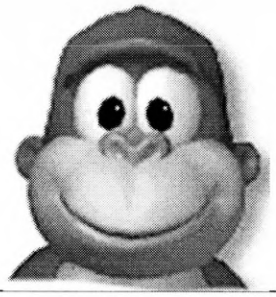
Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike



Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

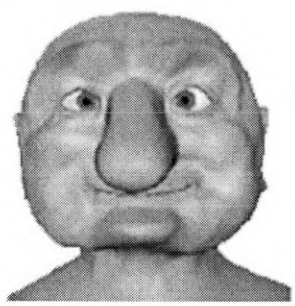
Appendix 9 continued



Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

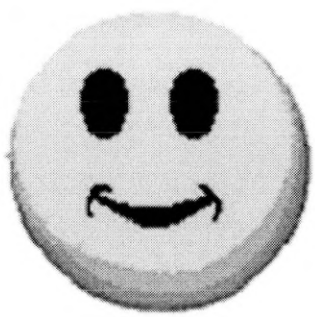
Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike



Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

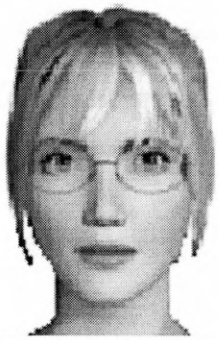


Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Appendix 9 continued

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

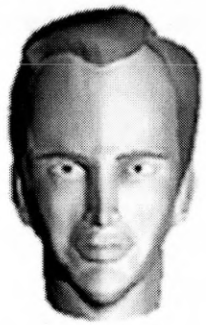


Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

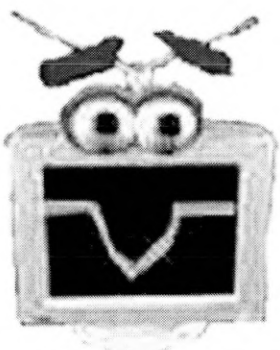
Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Appendix 9 continued



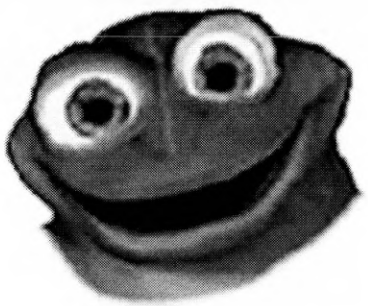
Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike



Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike



Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Unattractive	1	2	3	4	5	6	7	8	9	10	Attractive
Trustworthy	1	2	3	4	5	6	7	8	9	10	Untrustworthy
Sensible	1	2	3	4	5	6	7	8	9	10	Foolish
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Useful	1	2	3	4	5	6	7	8	9	10	Useless
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Intelligent	1	2	3	4	5	6	7	8	9	10	Stupid
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Like	1	2	3	4	5	6	7	8	9	10	Dislike

Sex: M F

Age: ____

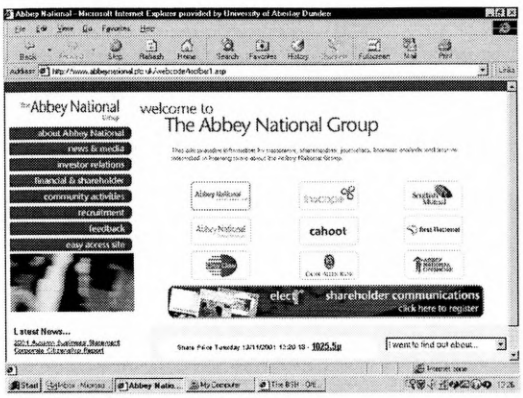
Please rate each character on all 9 dimensions by circling the appropriate number.
e.g.

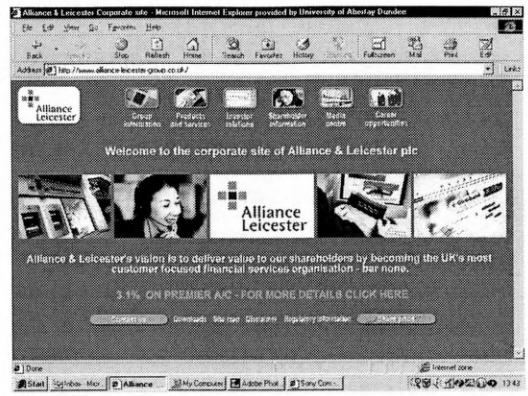
Unfriendly 1 2 (3) 4 5 6 7 8 9 10 Friendly

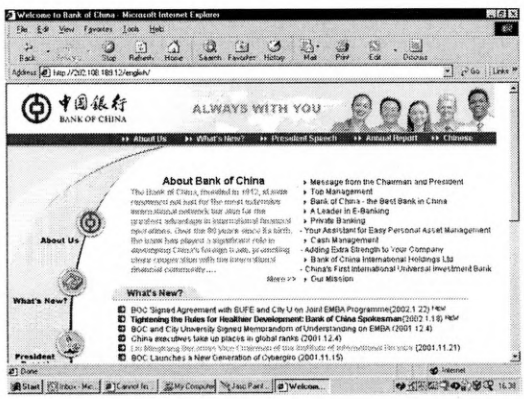
where 1 is unfriendly and 10 is friendly

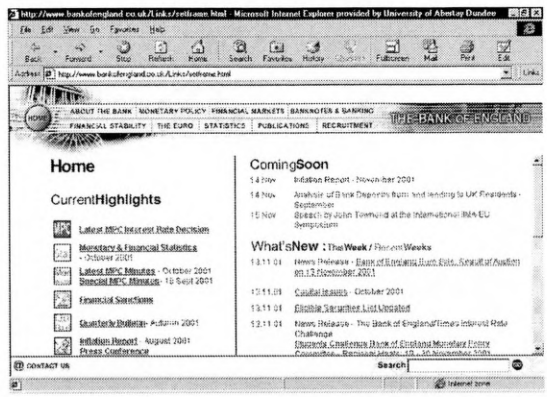
Appendix 10

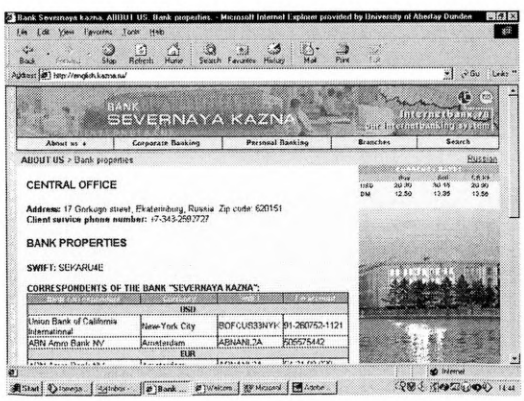
34 websites used in Chapter 5, phase 3

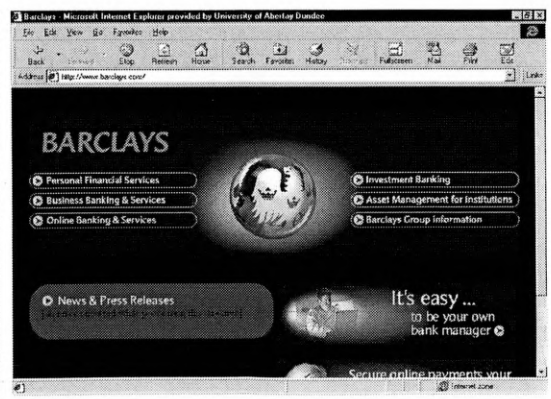


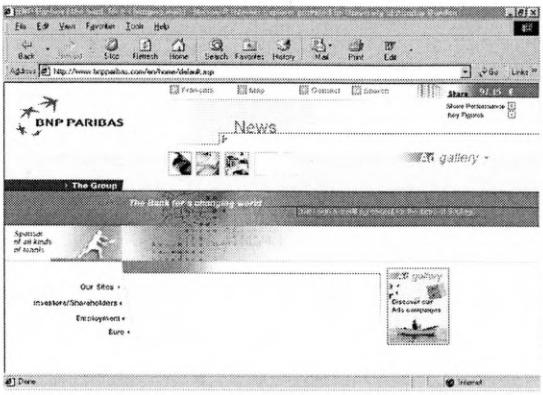


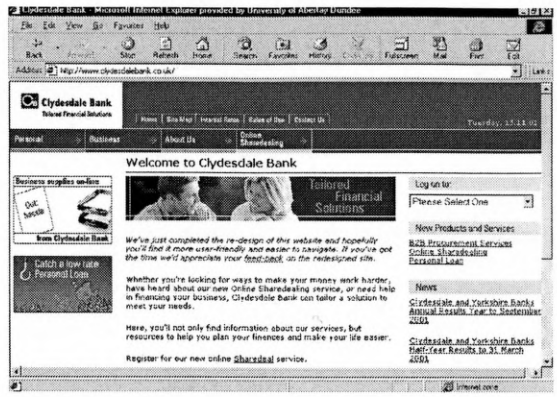


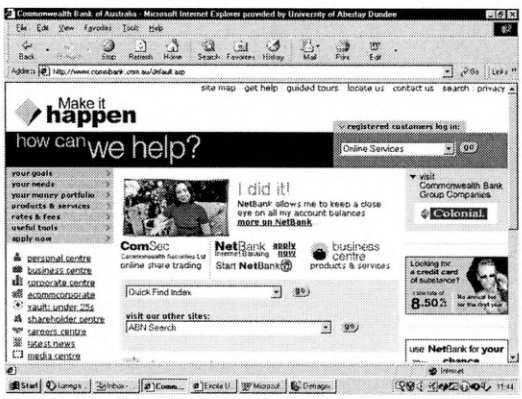






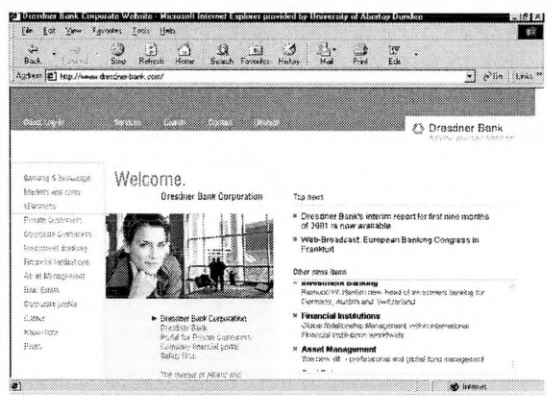

BNP PARIBAS
News
The Group


Clydesdale Bank
Welcome to Clydesdale Bank
Tailored Financial Solutions


Make it happen
how can we help?
I did it!


BANQUE POPULAIRE DE LA CÔTE D'AZUR
A team and product adapted to international needs

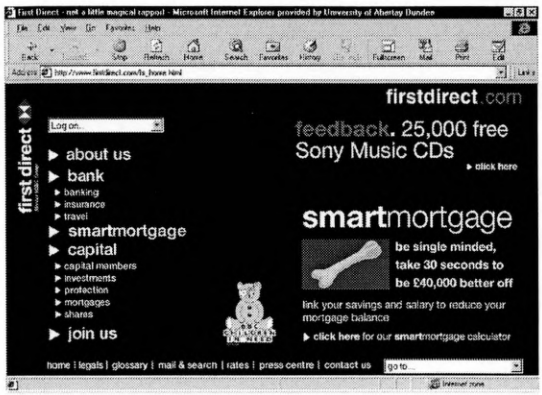

Deutsche Bank

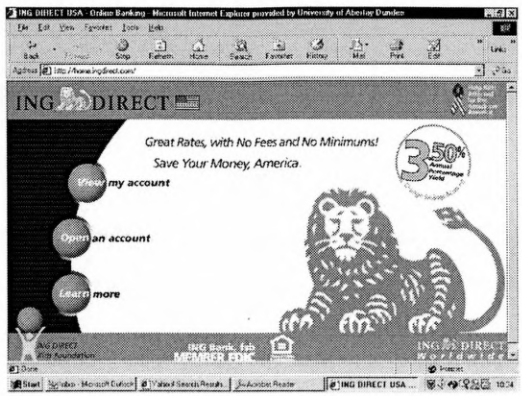

Dresdner Bank
Welcome.

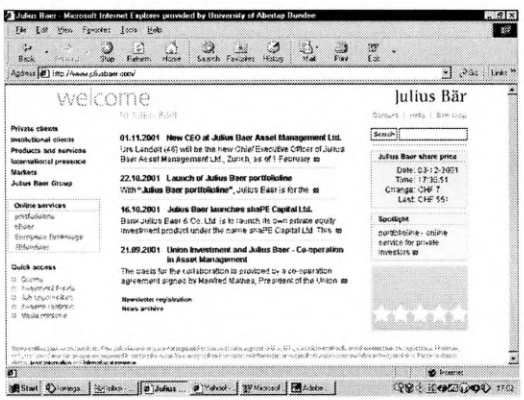

egg:
Service your accounts
Bank
Invest
Insure
Shop
Egg free Zone

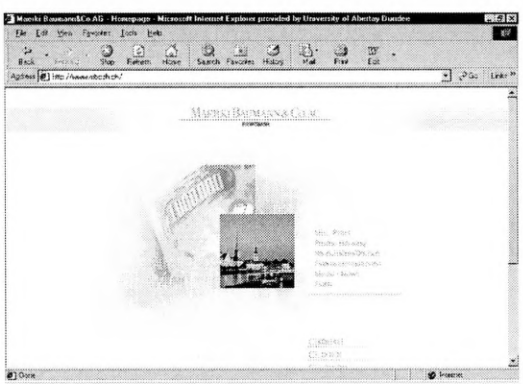

WELCOME TO
FAR EASTERN BANK

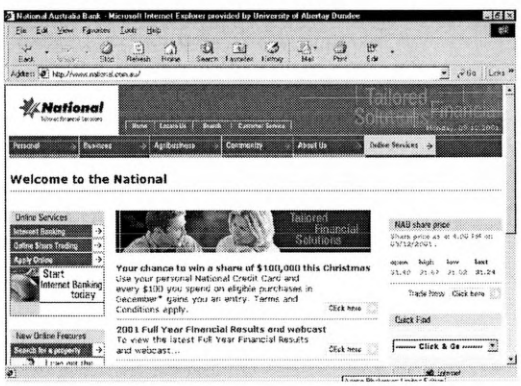
Appendix 10 continued

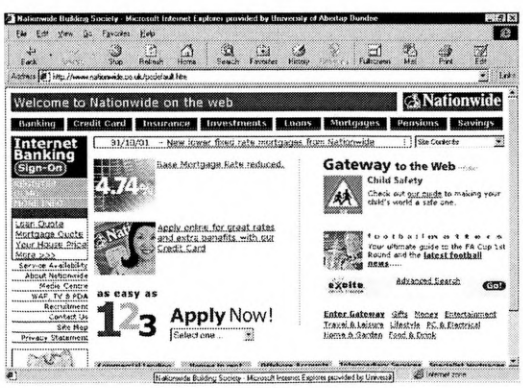


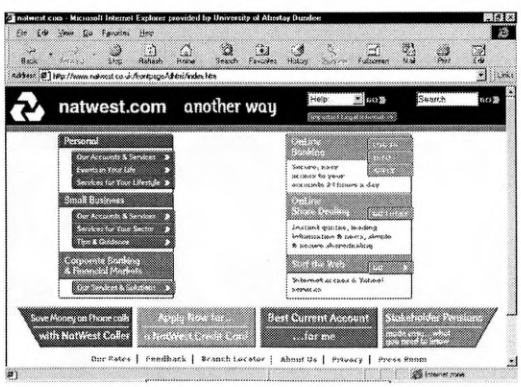


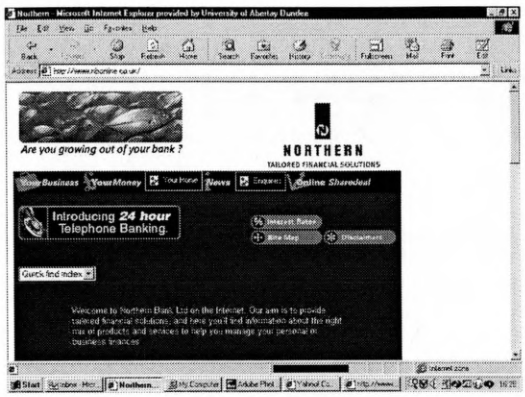


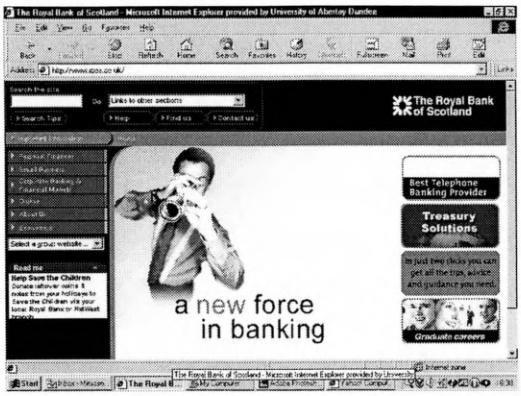


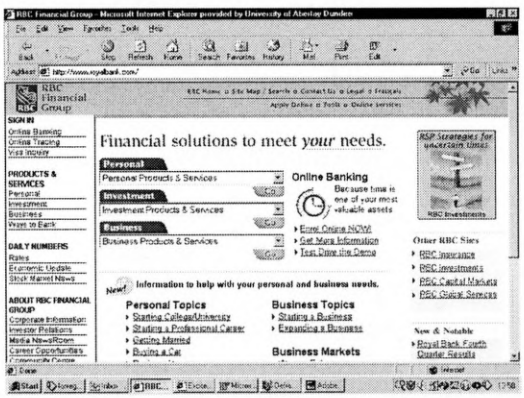







Are you growing out of your bank ?
NORTHERN
TAILORED FINANCIAL SOLUTIONS
Introducing 24 hour Telephone Banking.


The Royal Bank of Scotland
a new force in banking
Best Telephone Banking Provider
Treasury Solutions
Graduate careers


RBC Financial Group
Financial solutions to meet your needs.
Personal
Investment
Business
Online Banking
Personal Topics
Business Topics
Business Markets


Sainsbury's Bank
open for you
We are Sainsbury's Bank.
Our aim is to provide the best deals to make your life easier.
call us free on
0500
40 50 60†

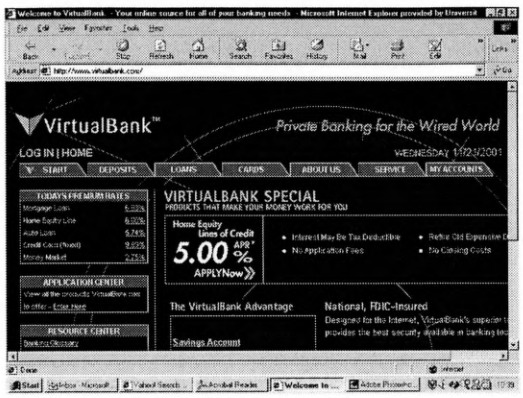

welcome to smile the internet bank
16th November 2001 'raise a smile' for Children in Need

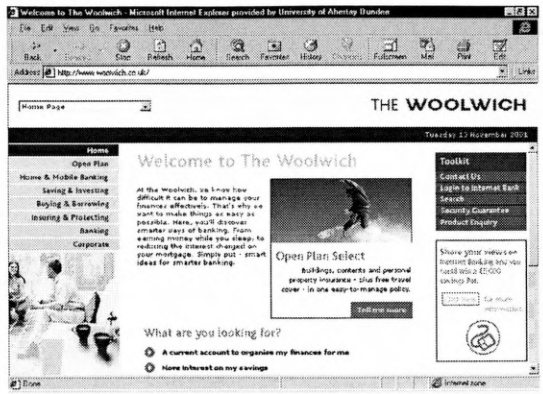

We've got the mortgage that's right for you.
The best mortgage statement ever?
Need help and advice from Alvin Hall and Which?


welcome to TESCO Personal Finance
for information & to apply click a product below
loans
credit card
savings
ISA
mortgage finder
life insurance
motor insurance
home insurance
pet insurance
travel insurance
travel money
finance offers!
click here for great value!
online banking


UNIBANCO
Internet 30 Horas
Personal Finance
To do good business...
Business Services
30 Horas
30 Hour Services

Appendix 10 continued





Appendix 11

Instructions and website questionnaire used in Chapter 5, phase 3

Instructions

You are going to be presented with a series of images of financial company's web sites. Your task is to rate each web site in turn on a number of dimensions and then to answer a few short questions on each questionnaire provided. Please indicate at the beginning of the questionnaire which web site you are rating. When you have finished rating a web site, press any key to display the next one.

Web Site _____

Please rate each of the web sites shown by circling a number on the scales below.

Ugly	1	2	3	4	5	6	7	8	9	10	Beautiful
Inappropriate	1	2	3	4	5	6	7	8	9	10	Appropriate
Forgettable	1	2	3	4	5	6	7	8	9	10	Memorable
Cold	1	2	3	4	5	6	7	8	9	10	Warm
Serious	1	2	3	4	5	6	7	8	9	10	Fun
Unappealing	1	2	3	4	5	6	7	8	9	10	Appealing
Untrustworthy	1	2	3	4	5	6	7	8	9	10	Trustworthy
Unfriendly	1	2	3	4	5	6	7	8	9	10	Friendly
Awkward	1	2	3	4	5	6	7	8	9	10	Elegant
Boring	1	2	3	4	5	6	7	8	9	10	Interesting
Unstructured	1	2	3	4	5	6	7	8	9	10	Structured
Unusable	1	2	3	4	5	6	7	8	9	10	Usable
Illogical layout	1	2	3	4	5	6	7	8	9	10	Logical layout
Chaotic	1	2	3	4	5	6	7	8	9	10	Calm
Unintuitive	1	2	3	4	5	6	7	8	9	10	Intuitive
Opaque	1	2	3	4	5	6	7	8	9	10	Transparent
Cluttered	1	2	3	4	5	6	7	8	9	10	Uncluttered
Tacky	1	2	3	4	5	6	7	8	9	10	Professional

What was your Overall Impression of the web site?

Bad	1	2	3	4	5	6	7	8	9	10	Good

Appendix 12

Means and standard deviations for each website, on each dimension

Website	beauty	appropriate	memorable	warm	fun	appeal	trust	friendly	elegant
1	6.00	6.89	5.11	5.56	4.89	5.67	6.67	5.78	6.44
	1.58	1.05	1.83	1.24	1.54	1.66	1.41	1.48	1.59
2	5.56	6.11	5.78	6.67	5.78	6.00	6.89	7.00	6.00
	2.01	1.76	1.72	1.41	1.20	2.00	1.27	1.58	0.71
3	3.44	6.67	2.89	3.11	3.00	2.56	5.89	2.67	4.11
	1.67	1.73	1.36	1.62	1.80	1.59	1.36	1.32	1.36
4	4.67	4.78	3.56	6.11	5.00	4.44	5.33	5.78	5.00
	2.06	2.77	1.67	1.90	1.87	1.88	1.41	1.92	1.66
5	5.00	6.22	6.00	5.11	5.89	5.44	6.11	6.33	5.44
	1.87	1.79	2.35	2.15	2.20	2.30	1.90	1.58	1.24
6	3.89	4.78	4.22	4.11	4.11	3.44	5.00	3.67	4.33
	1.76	1.20	2.44	1.76	2.03	2.24	1.32	1.58	1.12
7	5.56	5.89	4.67	4.33	3.33	5.11	5.33	4.00	5.11
	2.07	1.54	1.80	1.58	1.66	2.57	2.45	1.80	2.03
8	7.33	7.56	6.67	6.67	5.11	7.78	7.78	6.67	7.78
	1.50	1.24	1.87	1.80	1.54	1.56	1.09	1.22	1.48
9	7.11	5.44	6.11	6.78	6.44	6.78	6.00	7.11	6.67
	1.54	1.42	1.45	1.30	2.07	1.99	1.32	1.36	1.94
10	4.80	5.70	4.30	5.90	5.20	4.90	5.60	5.70	5.00
	1.48	1.25	1.64	2.08	1.32	1.37	1.43	1.83	1.63
11	4.44	5.22	5.00	5.22	5.00	4.22	5.22	5.33	4.33
	2.13	1.39	2.12	1.86	2.35	2.33	0.97	2.00	1.73
12	6.33	4.78	5.56	7.11	6.67	5.78	4.56	6.44	5.56
	2.06	1.99	2.24	1.27	1.22	2.33	1.67	1.42	1.94
13	5.89	6.56	5.33	4.56	5.00	5.78	6.11	5.44	5.44
	1.76	1.88	2.12	1.67	1.87	1.56	1.27	1.24	1.81
14	5.89	7.11	4.89	5.56	4.22	6.44	6.56	6.44	6.00
	2.09	1.05	2.32	2.60	1.72	1.51	1.24	1.81	1.94
15	5.33	5.78	5.33	6.44	6.56	5.67	5.67	7.00	5.00
	1.87	1.56	1.87	1.67	1.51	1.12	1.00	1.32	1.73
16	4.33	5.00	5.89	5.89	5.44	4.33	4.00	6.11	4.67
	1.41	1.87	2.03	1.27	1.33	2.24	2.06	1.90	1.73
17	4.33	5.22	6.56	4.44	4.78	5.00	5.44	5.33	4.89
	1.87	1.56	2.13	2.46	2.49	2.24	1.59	2.12	2.03
18	5.22	5.22	7.67	7.33	7.00	5.44	5.22	6.22	5.67
	2.05	1.92	1.22	0.87	1.00	1.81	1.72	1.72	1.87
19	4.00	4.67	3.00	2.89	2.44	3.00	5.56	3.22	4.56
	2.18	2.00	1.87	1.76	1.51	2.12	1.24	1.72	2.07
20	5.50	6.60	5.50	5.20	5.60	5.40	7.00	6.60	5.50
	1.65	1.07	1.72	1.75	1.07	2.07	1.05	1.35	1.72
21	7.80	6.40	6.80	6.00	4.90	7.10	6.60	6.60	7.60
	0.79	1.71	1.32	1.56	1.29	1.45	1.07	1.84	1.71
22	5.10	5.90	4.40	5.30	4.90	4.60	6.10	5.90	5.00
	1.29	1.60	1.84	1.34	1.52	1.71	0.99	1.45	1.56
23	3.90	4.40	5.40	6.50	6.20	4.30	5.10	6.50	4.00
	1.37	1.90	1.71	1.43	1.69	2.16	1.29	1.51	1.33

Appendix 12 continued

Website	interesting	structured	usable	logical layout	calm	intuitive	transparent	uncluttered	professional	overall impression
1	4.89	7.67	7.78	7.78	6.67	6.78	6.44	6.89	7.56	6.56
	1.90	1.12	1.64	1.48	1.22	1.30	1.13	1.90	1.24	1.42
2	6.00	7.44	7.78	7.11	5.89	6.56	6.22	6.89	7.67	6.67
	1.80	1.42	1.39	1.62	0.93	1.24	1.09	2.20	1.41	1.22
3	2.56	5.00	5.11	4.33	4.11	4.33	4.67	3.22	5.56	3.78
	1.42	2.35	1.27	2.24	1.69	1.80	2.00	2.49	2.24	1.09
4	5.00	3.89	5.22	4.00	3.33	4.33	5.11	3.11	4.67	4.56
	1.73	2.20	2.54	2.40	2.40	1.80	2.09	2.32	2.12	1.94
5	4.44	7.89	8.00	8.11	7.22	7.11	7.11	7.89	6.89	6.00
	2.35	1.62	1.58	1.27	1.99	1.69	1.45	1.36	2.42	1.87
6	3.44	4.44	4.89	4.33	4.56	4.67	4.67	3.78	4.44	4.11
	1.81	2.24	2.32	2.35	1.94	1.80	1.22	1.86	2.19	1.76
7	4.33	5.67	4.89	5.33	5.44	5.67	5.44	5.56	7.22	5.44
	2.24	1.41	1.96	1.94	2.19	2.06	1.81	2.19	1.30	2.07
8	6.89	8.33	8.33	8.11	8.11	7.67	7.44	8.56	8.78	7.78
	1.96	1.50	1.50	1.36	1.17	1.32	1.51	1.01	0.83	0.83
9	6.56	6.33	5.78	5.89	6.78	5.78	5.78	7.44	6.44	6.33
	1.88	1.94	1.39	1.76	1.92	1.86	1.48	2.19	2.35	2.00
10	4.70	5.20	5.70	5.00	4.80	4.70	5.10	4.40	6.00	5.20
	1.77	1.48	1.25	1.56	1.55	0.95	0.88	1.90	1.94	1.55
11	4.89	5.44	5.78	5.00	4.67	4.33	4.33	4.33	5.44	4.67
	3.14	2.07	1.72	2.18	2.74	1.50	1.66	2.40	1.51	1.94
12	6.33	4.44	5.22	5.00	6.22	4.89	5.00	6.00	4.11	4.89
	1.50	1.59	1.79	1.80	2.33	1.36	1.58	1.87	2.03	1.76
13	5.33	6.78	6.33	6.22	4.78	5.56	5.22	3.89	6.33	5.78
	2.00	2.05	1.22	1.39	1.56	1.51	1.56	2.62	2.24	1.86
14	5.78	7.22	7.00	7.00	6.44	6.00	6.44	6.22	7.78	6.56
	1.86	1.72	1.41	1.66	2.07	1.32	1.33	1.79	1.56	1.81
15	6.11	5.56	6.67	6.22	5.56	6.22	6.22	5.00	5.11	5.78
	1.45	1.33	1.32	1.64	1.42	1.48	1.09	2.12	1.62	1.39
16	3.89	6.89	6.67	6.44	6.33	5.56	5.78	7.11	4.00	4.67
	1.90	1.05	1.66	1.51	0.87	1.51	1.20	1.45	2.18	2.18
17	4.33	6.78	6.56	6.44	6.22	6.11	5.89	6.56	5.33	5.56
	1.80	1.79	1.51	1.59	1.86	1.69	1.62	2.07	2.35	2.01
18	6.22	6.67	7.00	7.44	7.00	6.89	6.78	8.00	5.44	6.11
	1.20	1.73	1.58	1.24	1.80	1.54	1.30	1.32	2.55	1.54
19	2.44	5.11	5.11	5.67	5.00	4.44	4.56	5.33	5.67	3.78
	1.88	2.15	2.03	1.66	2.29	2.07	1.59	2.35	3.12	1.92
20	5.40	6.60	6.50	6.30	5.70	6.00	5.60	5.60	6.20	5.80
	1.65	1.65	1.58	1.89	2.21	1.63	1.58	2.72	2.10	1.69
21	6.80	7.20	6.60	6.00	8.70	6.80	6.50	9.20	8.20	7.00
	1.69	1.87	1.65	1.70	1.49	1.40	1.90	0.79	1.62	2.00
22	5.20	6.00	5.60	6.10	5.20	5.90	5.10	5.30	5.60	5.30
	1.55	1.41	1.58	1.20	1.40	1.10	1.66	2.06	1.43	1.49
23	5.20	5.30	5.70	5.10	4.00	4.60	4.60	3.90	3.60	4.50
	1.75	2.00	1.64	1.37	1.56	1.17	1.17	2.02	1.51	1.27

Appendix 12 continued

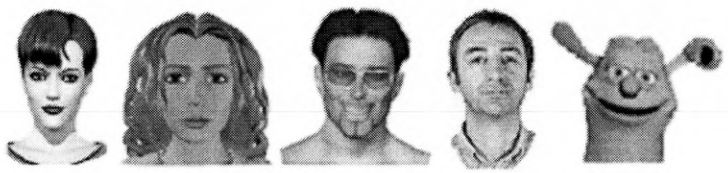
Website	beauty	appropriate	memorable	warm	fun	appeal	trust	friendly	elegant
24	5.20	6.70	5.30	4.80	4.00	4.70	6.80	5.30	6.10
	1.69	0.82	1.95	2.20	2.36	2.41	1.48	2.06	1.79
25	4.80	5.30	4.60	5.70	5.30	5.00	6.00	6.60	5.40
	1.55	1.16	2.01	1.49	1.42	1.89	1.33	1.58	1.51
26	4.20	5.10	3.50	4.40	3.90	3.30	5.20	4.20	4.00
	1.55	1.52	1.90	2.01	2.13	2.36	1.55	2.15	1.63
27	6.40	7.10	6.70	5.90	5.40	5.70	7.00	6.60	6.40
	1.26	1.37	1.49	0.99	1.17	2.06	1.05	1.07	1.43
28	6.50	6.10	7.10	8.30	8.10	7.60	6.60	8.50	6.70
	0.85	0.99	1.29	1.06	0.88	0.97	1.90	1.27	1.06
29	5.30	4.70	6.10	6.70	7.10	5.20	5.00	6.90	5.20
	1.16	2.11	1.91	1.64	2.13	1.81	1.33	1.45	1.75
30	5.00	5.40	5.30	5.80	5.40	4.10	5.50	5.70	4.90
	1.05	1.26	1.89	0.92	1.78	1.66	1.08	1.42	1.37
31	4.80	5.20	7.80	7.50	8.00	5.80	5.70	7.40	5.00
	2.35	1.99	1.99	1.84	1.05	2.10	1.95	2.17	1.83
32	5.90	7.00	5.50	5.20	3.80	5.30	6.70	5.30	5.90
	1.79	1.41	1.84	1.40	1.55	2.26	1.70	2.00	1.66
33	4.60	5.30	4.00	4.10	3.80	3.80	5.40	4.00	4.10
	2.32	1.89	1.89	1.91	1.99	2.30	1.35	1.83	1.97
34	5.40	6.50	4.80	5.10	4.70	5.40	6.50	5.80	5.90
	1.43	1.18	2.04	2.02	2.00	1.96	1.58	2.04	1.60

Website	interesting	structured	usable	logical layout	calm	intuitive	transparent	uncluttered	professional	overall impression
24	4.40	7.80	7.40	7.30	6.50	6.20	6.50	7.40	7.20	6.10
	2.32	1.62	1.17	0.95	1.08	1.48	1.27	1.17	1.69	1.73
25	4.70	5.40	6.00	6.00	5.70	5.20	5.20	6.10	5.70	5.70
26	3.30	4.80	4.70	4.30	2.80	4.00	4.30	2.50	4.90	3.40
	1.70	1.99	1.77	1.89	1.75	1.89	2.00	1.90	1.79	1.71
27	6.00	7.40	7.40	7.60	6.80	6.40	6.50	7.10	7.10	6.90
	1.15	0.97	1.07	0.70	0.79	0.84	1.08	1.45	1.79	0.88
28	7.60	7.80	8.10	7.70	7.80	7.30	7.00	8.30	5.60	7.50
	1.07	1.23	1.60	1.83	1.75	1.64	1.83	1.25	2.07	0.71
29	5.50	5.90	6.00	6.00	6.00	5.40	5.80	7.60	4.50	5.20
	1.58	1.45	1.49	2.05	1.56	1.78	1.87	2.12	2.12	1.48
30	5.50	4.70	5.20	4.80	3.60	4.40	4.70	3.60	4.70	4.20
	2.07	2.21	1.55	2.25	1.84	2.01	1.57	2.22	1.64	1.93
31	6.80	7.00	7.40	7.30	5.90	6.90	6.60	6.80	4.50	6.20
	1.48	2.00	2.12	1.83	2.08	1.97	2.07	2.04	2.80	2.10
32	5.10	6.70	6.70	6.80	5.90	5.80	6.60	5.90	6.70	6.00
	2.69	1.16	1.57	1.48	0.99	1.55	1.51	1.66	1.49	1.49
33	3.60	5.00	4.70	5.00	4.30	4.50	4.50	3.60	4.70	4.10
	2.41	2.75	2.21	2.16	2.11	2.27	2.12	2.27	2.91	2.08
34	5.20	7.10	6.90	7.10	6.50	6.50	6.80	5.50	6.80	5.90
	2.25	1.79	1.37	1.20	1.27	1.35	1.14	1.96	1.32	1.37

Appendix 13

Chapter 6 agents

Attractive agents

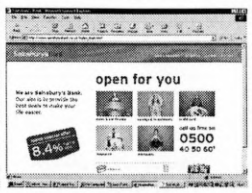


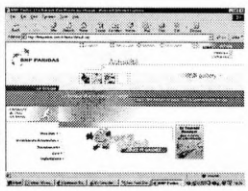
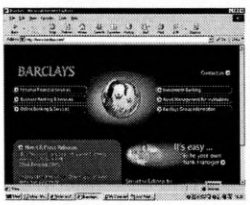
Agents of average attractiveness

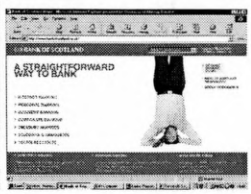
Unattractive agents

Appendix 14

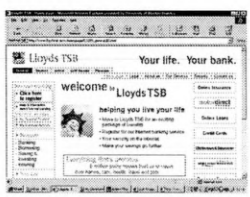
Chapter 6 websites

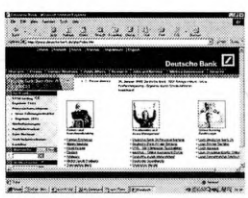



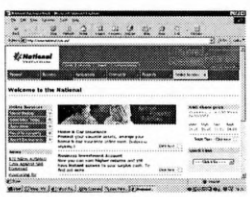



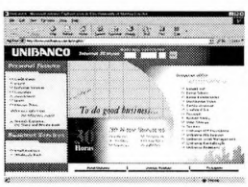


Good Websites

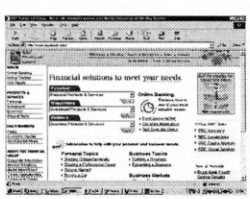


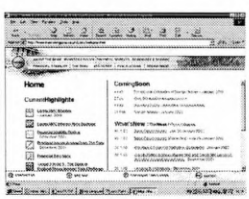


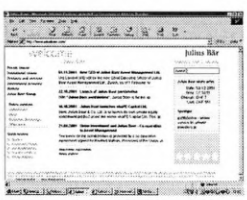




Average Websites







Bad Websites

Appendix 15

Chapter 6 agent questionnaire

Please rate the **agent** shown by circling a number shown on the scales below

Unattractive	1	2	3	4	5	6	7	Attractive
Memorable	1	2	3	4	5	6	7	Forgettable
Unappealing	1	2	3	4	5	6	7	Appealing
Untrustworthy	1	2	3	4	5	6	7	Trustworthy
Friendly	1	2	3	4	5	6	7	Unfriendly
Sensible	1	2	3	4	5	6	7	Foolish
Useless	1	2	3	4	5	6	7	Useful
Stupid	1	2	3	4	5	6	7	Intelligent
Like	1	2	3	4	5	6	7	Dislike

Appendix 16

Chapter 6 website questionnaire

Please rate each of the **web sites** shown by circling a number on the scales below.

Inappropriate	1	2	3	4	5	6	7	Appropriate
Forgettable	1	2	3	4	5	6	7	Memorable
Serious	1	2	3	4	5	6	7	Fun
Unappealing	1	2	3	4	5	6	7	Appealing
Untrustworthy	1	2	3	4	5	6	7	Trustworthy
Unfriendly	1	2	3	4	5	6	7	Friendly
Awkward	1	2	3	4	5	6	7	Elegant
Boring	1	2	3	4	5	6	7	Interesting
Unstructured	1	2	3	4	5	6	7	Structured
Unusable	1	2	3	4	5	6	7	Usable
Illogical layout	1	2	3	4	5	6	7	Logical layout
Chaotic	1	2	3	4	5	6	7	Calm
Unintuitive	1	2	3	4	5	6	7	Intuitive
Cluttered	1	2	3	4	5	6	7	Uncluttered
Tacky	1	2	3	4	5	6	7	Professional

What was your Overall Impression of the web site?

Bad	1	2	3	4	5	6	7	Good

Appendix 17

Mean Scores on each agent attribute, by agent and website condition

Mean attractiveness scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.50	4.66	4.43
Average	3.68	3.52	3.91
Unattractive	2.83	2.57	2.42

Mean memorability scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.20	4.46	4.37
Average	4.22	4.02	4.68
Unattractive	5.11	5.11	4.50

Mean appeal scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.68	4.46	4.37
Average	3.71	3.58	3.54
Unattractive	3.23	2.88	2.85

Mean trustworthiness scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.32	4.43	4.43
Average	3.80	3.68	3.28
Unattractive	3.28	3.37	3.12

Mean friendliness scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.95	4.57	4.28
Average	4.20	4.22	3.97
Unattractive	4.66	4.03	3.78

Mean sensibility scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.35	4.26	4.31
Average	3.80	3.70	3.20
Unattractive	3.20	3.46	3.08

Mean usefulness scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.48	4.66	4.28
Average	3.97	3.65	3.66
Unattractive	3.46	3.60	3.32

Mean intelligence scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.50	4.80	4.14
Average	3.80	3.90	3.71
Unattractive	3.68	3.28	3.32

Mean likeability scores

	Website Level		
Agent Level	Good	Average	Bad
Attractive	4.30	4.46	4.34
Average	3.71	3.40	3.83
Unattractive	3.77	3.66	2.85

Appendix 18

Mann-Whitney statistical summaries for differences between context conditions

Context v imagined

Test Statistics[a]

	attract	memorab	appeal	trust	friend	sensible	useful	intellig	like
Mann-Whitney U	30015.500	33491.000	30968.500	31275.000	34223.000	32573.000	30751.500	29744.000	30971.000
Wilcoxon W	84630.500	58916.000	85583.500	85890.000	88838.000	87188.000	85366.500	84359.000	85586.000
Z	-3.887	-1.992	-3.365	-3.200	-1.585	-2.485	-3.513	-4.048	-3.359
Asymp. Sig. (2-tailed)	.000	.046	.001	.001	.113	.013	.000	.000	.001

a. Grouping Variable: condition

Context v no context

Test Statistics[a]

	attract	memorab	appeal	trust	friend	sensible	useful	intellig	like
Mann-Whitney U	35214.000	36906.000	36476.000	35655.500	39564.000	38165.000	35607.500	35789.500	35263.500
Wilcoxon W	89829.000	91521.000	91091.000	90270.500	94179.000	92780.000	90222.500	90404.500	89878.500
Z	-2.292	-1.411	-1.632	-2.068	-.019	-.753	-2.123	-2.012	-2.268
Asymp. Sig. (2-tailed)	.022	.158	.103	.039	.985	.452	.034	.044	.023

a. Grouping Variable: condition

No context v imagined

Test Statistics[a]

	attract	memorab	appeal	trust	friend	sensible	useful	intellig	like
Mann-Whitney U	24215.000	22642.500	24407.500	24943.500	25066.000	24370.000	24415.500	23915.000	24790.000
Wilcoxon W	53135.000	48067.500	53327.500	53863.500	53986.000	53290.000	53335.500	52835.000	53710.000
Z	-1.952	-3.052	-1.815	-1.444	-1.353	-1.844	-1.834	-2.173	-1.547
Asymp. Sig. (2-tailed)	.051	.002	.070	.149	.176	.065	.067	.030	.122

a. Grouping Variable: condition

Appendix 19

Mean scores for male, female and cartoon agents, across all context conditions

Context

	Male Agents	Female Agents	Cartoon Agents
Attractive	2.95 (1.48)	4.29 (1.49)	3.52 (1.89)
Memorable	3.94 (1.78)	4.46 (1.41)	5.04 (1.57)
Appeal	3.18 (1.56)	4.18 (1.51)	3.98 (1.78)
Trust	3.38 (1.54)	4.45 (1.40)	3.41 (1.81)
Friendly	4.00 (1.69)	4.34 (1.50)	4.69 (2.14)
Sensible	3.72 (1.54)	4.29 (1.54)	2.79 (1.53)
Useful	3.70 (1.38)	4.45 (1.45)	3.38 (1.47)
Intelligent	3.82 (1.50)	4.51 (1.29)	3.23 (1.45)
Like	3.30 (1.52)	4.04 (1.44)	4.33 (1.98)

Imagined

	Male Agents	Female Agents	Cartoon Agents
Attractive	3.34 (1.88)	4.85 (1.67)	3.79 (2.21)
Memorable	4.02 (1.86)	4.01 (1.74)	5.38 (1.72)
Appeal	3.30 (1.86)	4.74 (1.61)	4.31 (2.22)
Trust	3.59 (1.73)	4.86 (1.39)	4.04 (1.98)
Friendly	4.18 (1.84)	4.50 (1.72)	5.10 (2.14)
Sensible	4.08 (1.72)	4.68 (1.75)	2.52 (1.83)
Useful	3.98 (1.58)	4.88 (1.39)	3.48 (1.66)
Intelligent	4.36 (1.70)	4.99 (1.46)	3.25 (1.68)
Like	3.45 (1.99)	4.66 (1.61)	4.69 (2.12)

No Context

	Male Agents	Female Agents	Cartoon Agents
Attractive	3.50 (1.68)	4.73 (1.35)	3.23 (1.67)
Memorable	4.18 (1.81)	4.53 (1.48)	6.14 (1.15)
Appeal	3.35 (1.65)	4.49 (1.35)	4.04 (2.26)
Trust	3.88 (1.40)	4.53 (1.22)	3.46 (1.60)
Friendly	4.00 (1.56)	4.32 (1.48)	5.06 (1.79)
Sensible	4.17 (1.23)	4.20 (1.27)	2.39 (1.16)
Useful	4.21 (1.19)	4.61 (1.13)	3.08 (1.38)
Intelligent	4.27 (1.16)	4.75 (1.18)	2.94 (1.46)
Like	3.71 (1.49)	4.38 (1.34)	4.33 (2.08)

Appendix 20

Mann-Whitney summary of male, female and cartoon agent differences between context and imagined context conditions

Male Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE	SUITS
Mann-Whitney U	4158.500	4484.000	4534.000	4301.000	4304.500	4043.500	4079.500	3734.000	4602.500	4385.500
Wilcoxon W	8814.500	9140.000	9190.000	8957.000	8960.500	8699.500	8735.500	8390.000	9258.500	9041.500
Z	-1.188	-.326	-.196	-.810	-.799	-1.488	-1.413	-2.309	-.015	-.735
Asymp. Sig. (2-tailed)	.235	.744	.845	.418	.424	.137	.158	.021	.988	.463

a. Grouping Variable: condition

b. agent gender = male

Female Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE	SUITS
Mann-Whitney U	3509.500	3902.000	3583.000	3945.000	4344.500	3931.000	3889.500	3685.500	3496.500	4228.000
Wilcoxon W	8165.500	8558.000	8239.000	8601.000	9000.500	8587.000	8545.500	8341.500	8152.500	8884.000
Z	-2.933	-1.867	-2.723	-1.768	-.695	-1.788	-1.911	-2.458	-2.944	-1.134
Asymp. Sig. (2-tailed)	.003	.062	.006	.077	.487	.074	.056	.014	.003	.257

a. Grouping Variable: condition

b. agent gender = female

Cartoon Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE	SUITS
Mann-Whitney U	1062.000	955.500	991.500	949.500	966.500	952.000	1084.000	1127.500	1021.000	1062.500
Wilcoxon W	2238.000	2131.500	2167.500	2125.500	2142.500	2128.000	2260.000	2303.500	2197.000	2238.500
Z	-.670	-1.482	-1.193	-1.509	-1.393	-1.509	-.508	-.183	-.873	-.929
Asymp. Sig. (2-tailed)	.503	.138	.233	.131	.164	.131	.611	.855	.331	.353

a. Grouping Variable: condition

b. agent gender = cartoon

Appendix 21

Mann-Whitney summary of male, female and cartoon agent differences between context and no context conditions

Male Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE
Mann-Whitney U	3730.500	4274.500	4311.000	3672.000	4588.500	3911.500	3705.000	3907.500	3849.500
Wilcoxon W	8386.500	8930.500	8967.000	8328.000	9244.500	8567.500	8361.000	8563.500	8505.500
Z	-2.321	-.877	-.785	-2.475	-.052	-1.859	-2.451	-1.880	-2.006
Asymp. Sig. (2-tailed)	.020	.380	.432	.013	.959	.063	.014	.060	.045

a. Grouping Variable: condition

b. agent gender = male

Female Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE
Mann-Whitney U	3905.500	4533.000	4176.000	4543.000	4445.500	4342.500	4460.500	4314.500	4098.000
Wilcoxon W	8561.500	9189.000	8832.000	9199.000	9101.500	8998.500	9116.500	8970.500	8754.000
Z	-1.875	-.199	-1.154	-.174	-.430	-.707	-.396	-.789	-1.360
Asymp. Sig. (2-tailed)	.061	.842	.248	.862	.667	.480	.692	.430	.174

a. Grouping Variable: condition

b. agent gender = female

Cartoon Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE
Mann-Whitney U	1058.500	615.000	1101.500	1127.000	1067.000	1005.500	1014.500	1009.500	1143.000
Wilcoxon W	2234.500	1791.000	2277.500	2303.000	2243.000	2181.500	2190.500	2185.500	2319.000
Z	-.695	-4.092	-.375	-.186	-.635	-1.115	-1.033	-1.067	-.067
Asymp. Sig. (2-tailed)	.487	.000	.708	.853	.525	.265	.301	.286	.947

a. Grouping Variable: condition

b. agent gender = cartoon

Appendix 22

Mann-Whitney summary of male, female and cartoon agent differences between imagined and no context conditions

Male Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE
Mann-Whitney U	4340.500	4394.000	4468.000	4122.500	4255.500	4557.000	4327.000	4302.500	4083.000
Wilcoxon W	8996.500	9050.000	9124.000	8778.500	8911.500	9213.000	8983.000	8958.500	8739.000
Z	-.704	-.564	-.369	-1.284	-.929	-.136	-.754	-.814	-1.383
Asymp. Sig. (2-tailed)	.481	.573	.712	.199	.353	.892	.451	.415	.167

a. Grouping Variable: condition

b. agent gender = male

Female Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE
Mann-Whitney U	4103.000	3807.500	3948.500	3912.500	4281.500	3666.500	3992.000	3991.000	3962.000
Wilcoxon W	8759.000	8463.500	8604.500	8568.500	8937.500	8322.500	8648.000	8647.000	8618.000
Z	-1.350	-2.111	-1.756	-1.858	-.862	-2.494	-1.656	-1.648	-1.714
Asymp. Sig. (2-tailed)	.177	.035	.079	.063	.389	.013	.098	.099	.086

a. Grouping Variable: condition

b. agent gender = female

Cartoon Agents

Test Statistics[a,b]

	ATTRACT	MEMORAB	APPEAL	TRUST	FRIEND	SENSIBLE	USEFUL	INTELLIG	LIKE
Mann-Whitney U	962.000	842.500	1093.500	934.000	1071.000	1062.000	940.500	1041.500	1022.500
Wilcoxon W	2138.000	2018.500	2269.500	2110.000	2247.000	2238.000	2116.500	2217.500	2198.500
Z	-1.415	-2.383	-.437	-1.617	-.610	-.683	-1.584	-.827	-.963
Asymp. Sig. (2-tailed)	.157	.017	.662	.106	.542	.495	.113	.408	.335

a. Grouping Variable: condition

b. agent gender = cartoon

Appendix 23

Mean scores for male, female and cartoon agents, across all context conditions

Attractive Agents

Attractive	4.85 (1.39)
Memorable	4.28 (1.71)
Appeal	4.70 (1.55)
Trust	4.67 (1.47)
Friendly	4.73 (1.65)
Sensible	4.45 (1.69)
Useful	4.65 (1.40)
Intelligent	4.69 (1.44)
Like	4.64 (1.58)

Agents of average attractiveness

Attractive	4.14 (1.56)
Memorable	4.26 (1.78)
Appeal	4.03 (1.68)
Trust	3.83 (1.47)
Friendly	4.24 (1.68)
Sensible	3.76 (1.53)
Useful	3.98 (1.39)
Intelligent	4.09 (1.41)
Like	4.01 (1.60)

Unattractive agents

Attractive	2.63 (1.60)
Memorable	4.90 (1.67)
Appeal	3.02 (1.67)
Trust	3.42 (1.66)
Friendly	4.14 (1.92)
Sensible	3.34 (1.61)
Useful	3.63 (1.43)
Intelligent	3.68 (1.59)
Like	3.46 (1.78)

Appendix 24

Mann-Whitney summary of attractive, average and unattractive agent differences between context and imagined conditions

Test Statistics[a]

agent_attractiveness		attract	memorab	appeal	trust	friend	sensible	useful	intellig	like
Attractive	Mann-Whitney U	2881.000	3640.000	3271.000	2809.500	3513.000	3126.500	3146.000	3000.000	2946.000
	Wilcoxon W	8986.000	6490.000	9376.000	8914.500	9618.000	9231.500	9251.000	9105.000	9051.000
	Z	-3.571	-1.376	-2.445	-3.758	-1.739	-2.833	-2.809	-3.219	-3.347
	Asymp. Sig. (2-tailed)	.000	.169	.014	.000	.082	.005	.005	.001	.001
Average	Mann-Whitney U	2669.500	4043.000	2965.500	3365.500	3572.000	3227.000	3192.500	3119.000	3082.500
	Wilcoxon W	8774.500	6893.000	9070.500	9470.500	9677.000	9332.000	9297.500	9224.000	9187.500
	Z	-4.161	-.233	-3.293	-2.164	-1.569	-2.553	-2.672	-2.872	-2.959
	Asymp. Sig. (2-tailed)	.000	.815	.001	.030	.117	.011	.008	.004	.003
Unattractive	Mann-Whitney U	4111.000	3454.500	4026.000	4083.500	3982.000	3995.000	3788.000	3574.000	4122.000
	Wilcoxon W	10216.000	6304.500	10131.000	10188.500	6832.000	6845.000	9893.000	9679.000	6972.000
	Z	-.040	-1.913	-.282	-.118	-.405	-.389	-.966	-1.566	-.009
	Asymp. Sig. (2-tailed)	.968	.056	.778	.906	.685	.712	.334	.117	.993

a. Grouping Variable: condition

Appendix 25

Mann-Whitney summary of attractive, average and unattractive agent differences between context and no context conditions

Test Statistics[a]

agent_attractiveness		attract	memorab	appeal	trust	friend	sensible	useful	intellig	like
Attractive	Mann-Whitney U	3641.500	4236.000	4224.000	4184.500	4259.500	4219.500	4324.000	4200.500	4087.500
	Wilcoxon W	9746.500	10341.000	10329.000	10289.500	7499.500	7459.500	10429.000	10305.500	10192.500
	Z	-2.083	-.445	-.481	-.644	-.382	-.491	-.209	-.548	-.851
	Asymp. Sig. (2-tailed)	.037	.656	.630	.520	.703	.623	.835	.584	.395
Average	Mann-Whitney U	3566.000	4354.500	3511.000	3966.000	4351.500	4309.000	3982.500	3893.000	3515.500
	Wilcoxon W	9671.000	7594.500	9616.000	10071.000	7591.500	10414.000	10087.500	9998.000	9620.500
	Z	-2.273	-.123	-2.412	-1.182	-.132	-.249	-1.168	-1.397	-2.404
	Asymp. Sig. (2-tailed)	.023	.902	.016	.237	.895	.803	.243	.162	.016
Unattractive	Mann-Whitney U	4305.500	3618.000	4284.500	3684.000	4255.500	3714.500	3571.000	3728.000	4136.000
	Wilcoxon W	10410.500	9723.000	7534.500	9789.000	10360.500	9819.500	9676.000	9833.000	10241.000
	Z	-.259	-2.144	-.287	-1.947	-.391	-1.868	-2.289	-1.836	-.715
	Asymp. Sig. (2-tailed)	.795	.032	.774	.052	.696	.062	.022	.066	.474

a. Grouping Variable: condition

Appendix 26

Mann-Whitney summary of attractive, average and unattractive agent differences between no context and imagined conditions

Test Statistics[a]

agent_attractiveness		attract	memorab	appeal	trust	friend	sensible	useful	intellig	like
Attractive	Mann-Whitney U	2528.500	2524.500	2471.500	2196.000	2483.000	2179.500	2288.500	2319.000	2270.000
	Wilcoxon W	5788.500	5374.500	5711.500	5436.000	5723.000	5419.500	5528.500	5559.000	5510.000
	Z	-1.748	-1.727	-1.941	-2.940	-1.882	-2.978	-2.612	-2.487	-2.659
	Asymp. Sig. (2-tailed)	.080	.084	.052	.003	.060	.003	.009	.013	.008
Average	Mann-Whitney U	2379.000	2914.500	2718.500	2847.000	2589.000	2316.000	2505.000	2589.500	2731.500
	Wilcoxon W	5619.000	5764.500	5958.500	5887.000	5829.000	5556.000	5745.000	5829.500	5971.500
	Z	-2.275	-.310	-1.024	-1.301	-1.492	-2.504	-1.832	-1.514	-.979
	Asymp. Sig. (2-tailed)	.023	.756	.306	.193	.136	.012	.067	.130	.328
Unattractive	Mann-Whitney U	2984.000	2023.500	2856.000	2600.500	2757.500	2434.500	2752.500	2991.500	2864.000
	Wilcoxon W	5834.000	4873.500	6096.000	5450.500	5607.500	5284.500	5602.500	5841.500	5714.000
	Z	-.059	-3.571	-.526	-1.452	-.878	-2.065	-.916	-.031	-.494
	Asymp. Sig. (2-tailed)	.953	.000	.599	.146	.380	.039	.360	.975	.621

a. Grouping Variable: condition

Appendix 27

List of occupations in Chapter 7, study 1

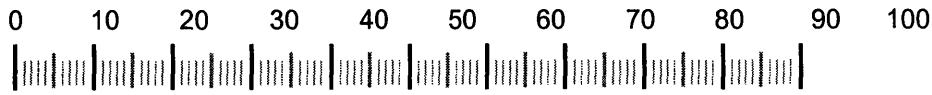
Bank teller
Bank manager
Counsellor
Doctor
Estate agent
Financial advisor
Fitness instructor
Florist
Holiday rep
Lawyer
Mechanic
Nurse
Shop assistant
Weather forecaster

Appendix 28

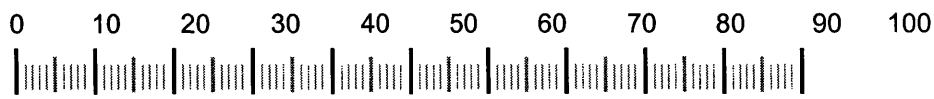
Occupation questionnaire in Chapter 7, study 1

Occupation: **Bank manager**

1. What % of bank managers do you think are female? Please mark on the scale below with a line.



2. What % of bank managers do you think are male? Please mark on the scale below with a line.



3. How well paid do you think a bank manager is?

1	2	3	4	5	6	7
Poorly paid						Highly paid

4. Over what age ranges do you think the bulk of the people employed as bank managers are? **You may circle more than one age range**. For example, if you think the job mainly employed people in the ages 16 to 35 then circle 16 – 25 and 26 – 35.

16 - 25 26 – 35 36 – 45 46 – 55 56+

5. How important do you think it is to be able to trust a bank manager?

1	2	3	4	5	6	7
Not important at all						Very important

6. How important is the level of intelligence of a person employed as a bank manager?

1	2	3	4	5	6	7
Not important at all						Very Important

7. How important is quality of advice from a bank manager?

1	2	3	4	5	6	7
Not important at all						Very Important

8. How important is it for a bank manager to be supportive and empathic?

1	2	3	4	5	6	7
Not important at all						Very Important

9. How important do you think it is for a bank manager to be physically attractive?

1	2	3	4	5	6	7
Not important at all						Very Important

10. How prestigious do you think the job of a bank manager is?

1	2	3	4	5	6	7
Not prestigious at all						Very Prestigious

11. Do you prefer your bank manager to be intelligent?

1	2	3	4	5	6	7
Not important at all						Very Important

12. Do you prefer your bank manager to be supportive and empathic?

1	2	3	4	5	6	7
Not important at all						Very Important

13. Do you prefer your bank manager to be physically attractive?

1	2	3	4	5	6	7
Not important at all						Very Important

Appendix 29

Mann-Whitney summary of difference between male and female occupations

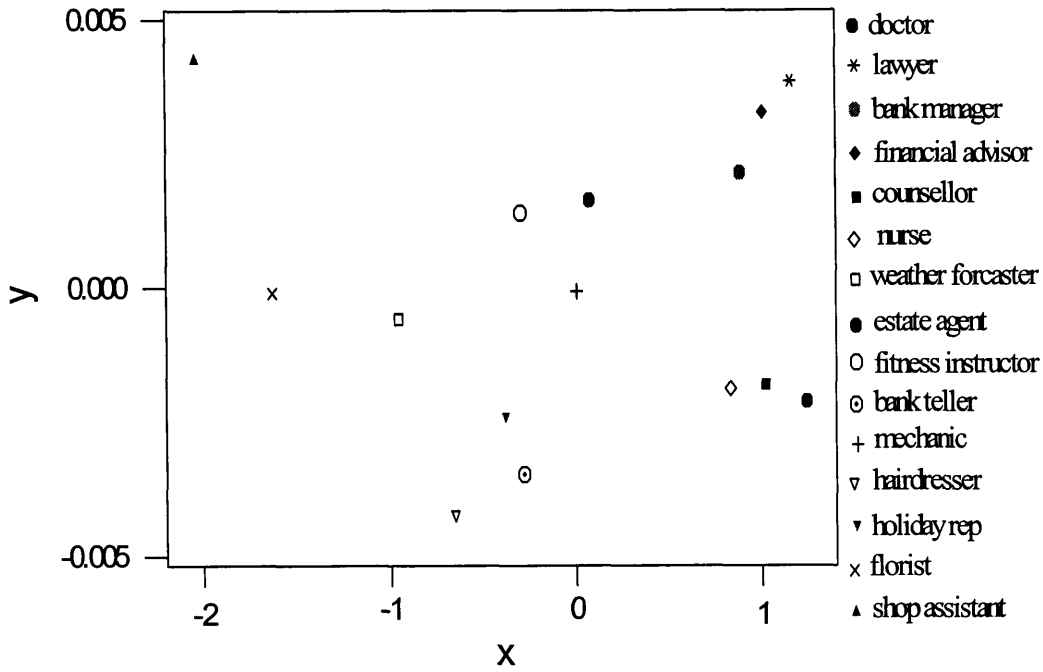
Test Statistics[a]

	well paid	TRUST	intelligence	ADVICE	SUPPORT	ATTRACT	prestigious	preference intelligence	preference supportive	preference attractive
Mann-Whitney U	4825.000	7277.500	5918.500	7130.500	9227.000	7273.500	7430.000	7138.500	10341.000	8996.500
Wilcoxon W	20578.000	23030.500	21671.500	22883.500	16248.000	14294.500	23183.000	22891.500	17362.000	16017.500
Z	-8.016	-4.700	-6.403	-4.788	-1.724	-4.632	-4.269	-4.681	-.145	-2.098
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.085	.000	.000	.000	.885	.036

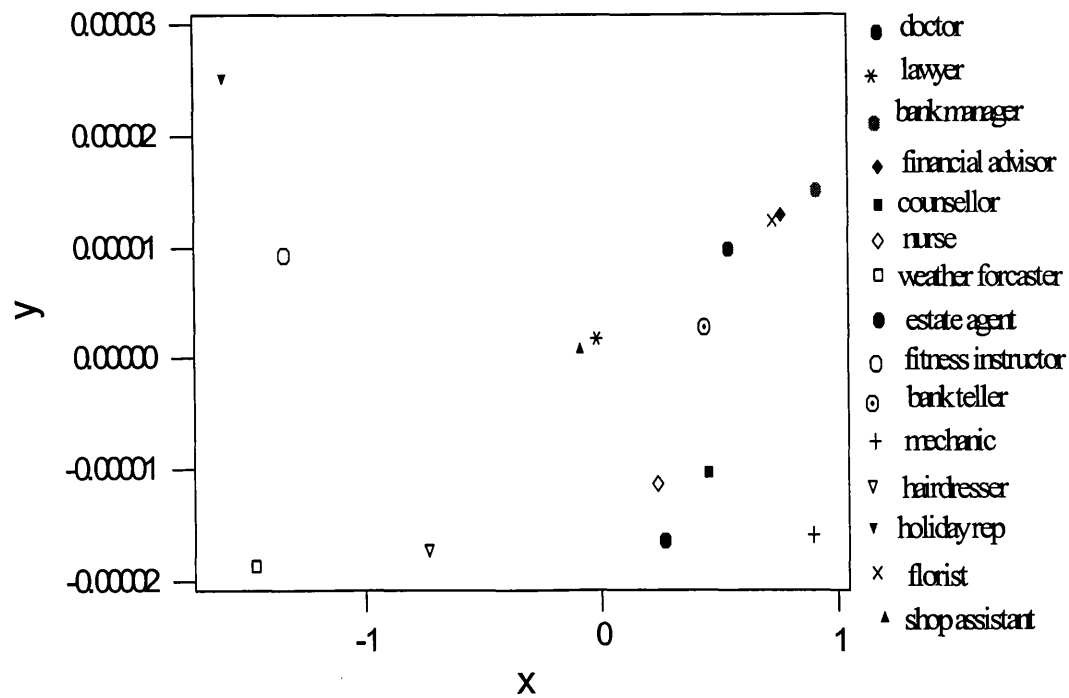
a. Grouping Variable: occupational gender based on results

Appendix 30

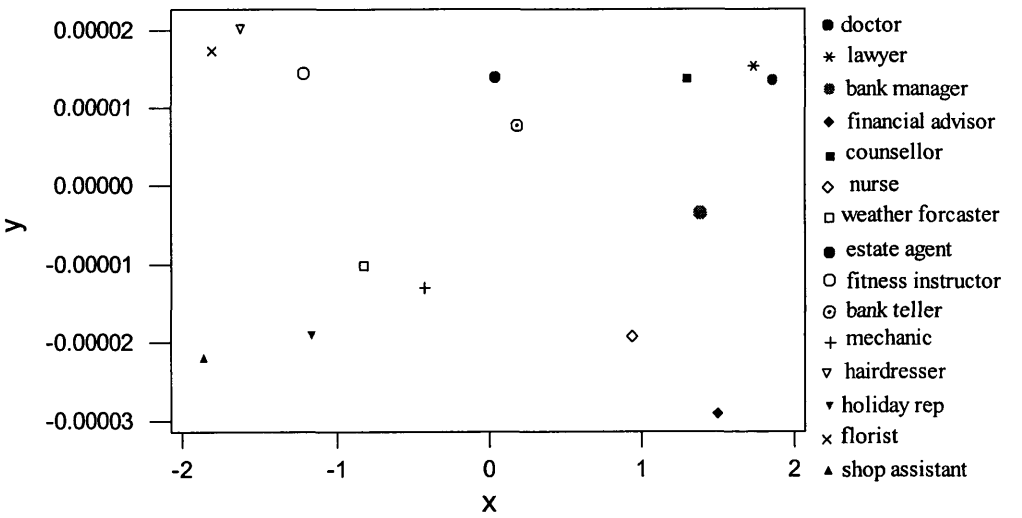
Multi-dimensional Scaling plots for each attribute measured in Chapter 7, study 1



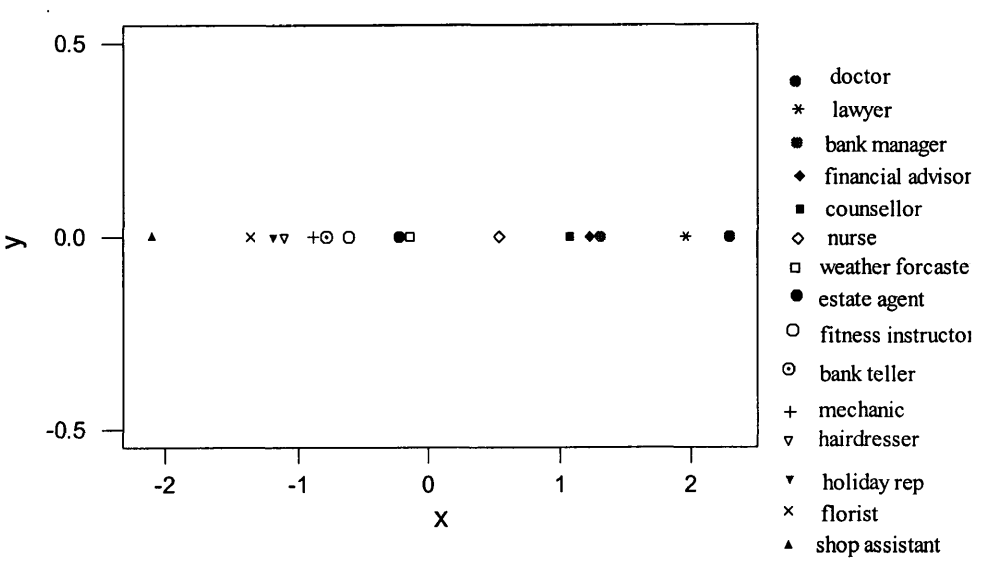
MDS Plot of Advice



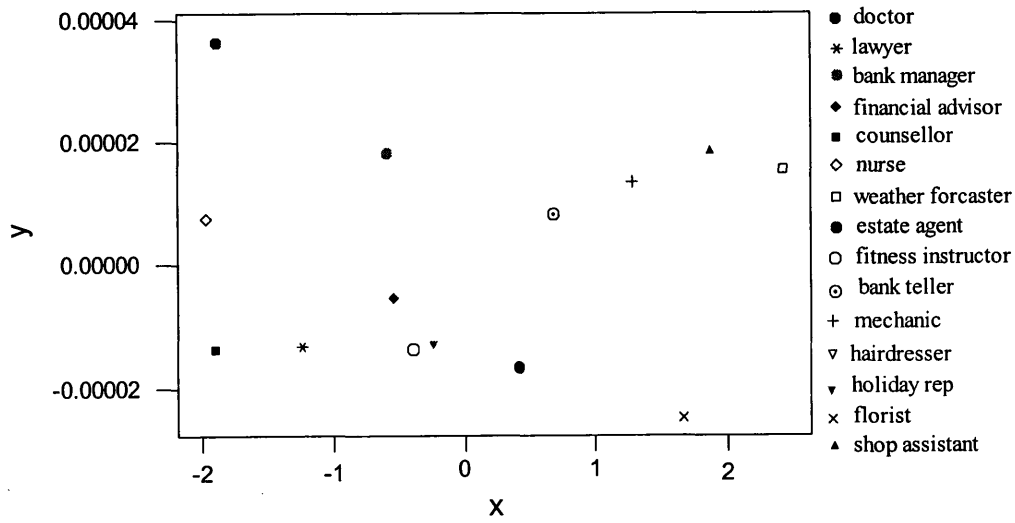
MDS Plot of Attractiveness



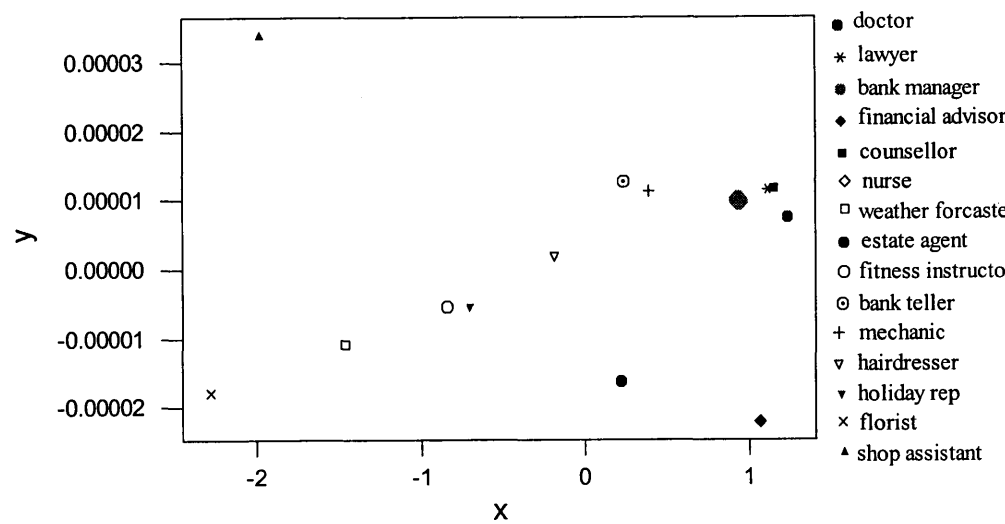
MDS Plot of Intelligence



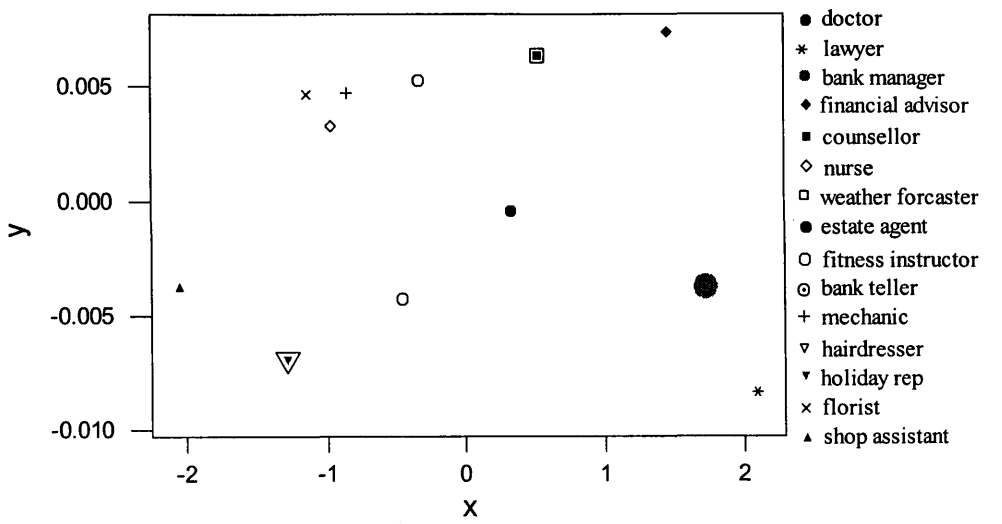
MDS Plot of Prestige



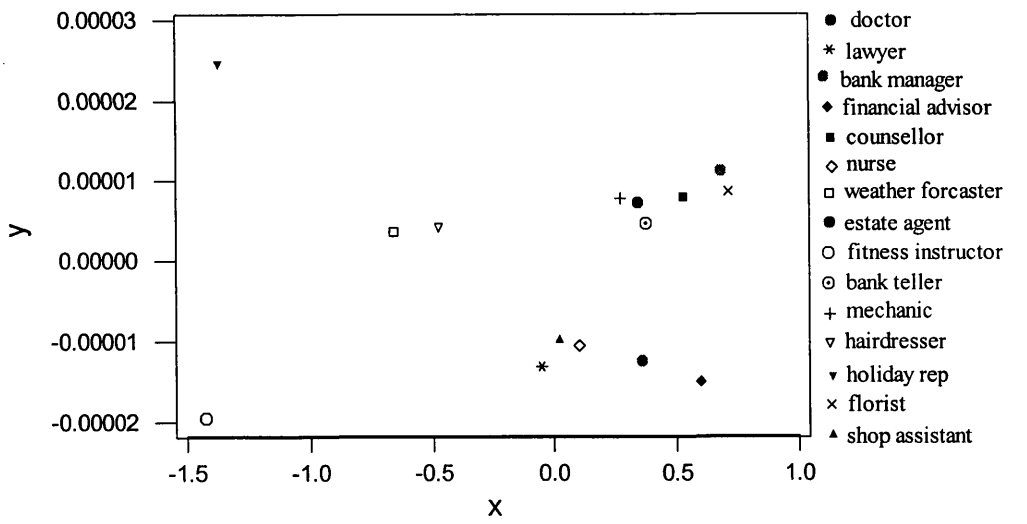
MDS Plot of Support



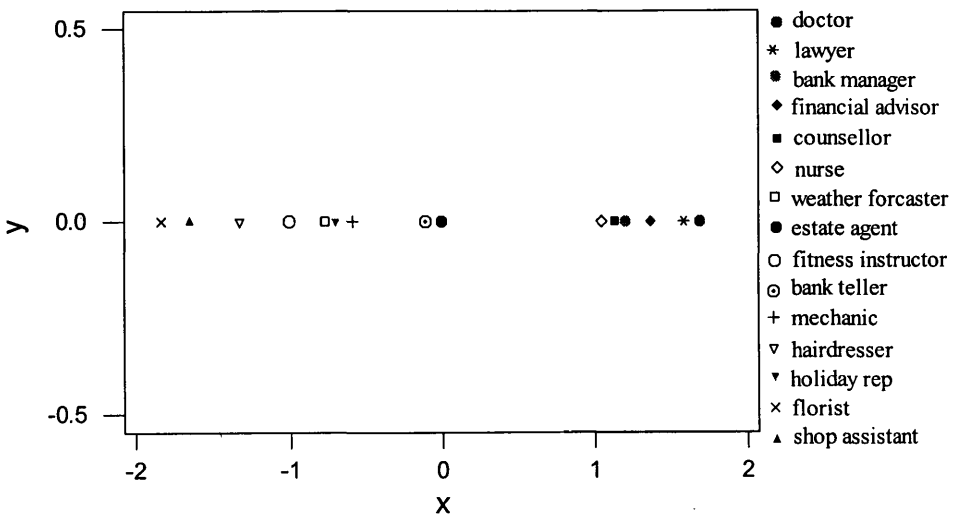
MDS Plot of Trust



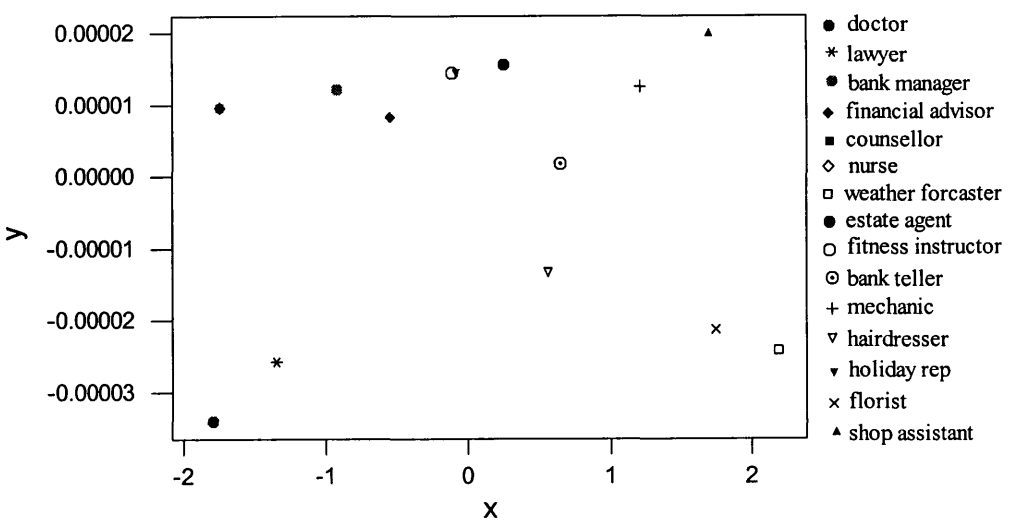
MDS Plot of Well Paid



MDS Plot of Preferred Attractiveness



MDS Plot of Preferred Intelligence



MDS Plot of Preferred Support

Appendix 31

List of occupations in Chapter 7, study 2 (pilot phase) and rating scales

Radiologist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Chemical Engineer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Chef
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Computer Programmer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Fashion Designer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Firefighter
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Funeral Director
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Architect
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Geologist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Bus Driver
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Labourer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Hotel Porter
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Carpenter
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Soldier
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Miner
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Welder
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Farmer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Pilot
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Vet
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Cabin Crew/Flight Attendant
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Paramedic
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Shop Assistant
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Mechanic
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Social Worker
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Legal Secretary
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

High School Teacher
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Florist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Dentist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Journalist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Car Dealer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Art Critic
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Driving Instructor
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Bank Manager
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Company Director
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Secretary
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Paediatrician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Prison Guard
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Taxi Driver
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Builder
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Psychiatrist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Lifeguard
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Midwife
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Judge
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Dietician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Photographer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Graphic Designer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Customer Service Advisor
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Beautician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Civil Servant
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Traffic Warden
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

University Lecturer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Receptionist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Nursery teacher
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Librarian
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Flight Engineer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Electrician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Painter/Decorator
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Police Officer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Factory Machine Operator
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Butcher
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Gynaecologist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Market Researcher
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Politician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Stockbroker
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Taxidermist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Surveyor
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Gardener
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Technician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Jeweller
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Real Estate Agent
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Professor
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Statistician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Lawyer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Novelist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Minister/Priest
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Bartender
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Auctioneer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Union Official
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Food Store Manager
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Physician
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Child care worker
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Hotel Manager
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Surgeon
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Web Designer
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Accountant
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Private Investigator
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Cashier
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Dental Hygienist
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Housekeeper
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Waiter
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

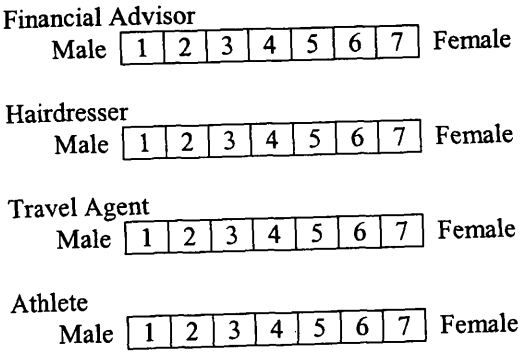
Barber
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Fitness Instructor
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Bank Teller
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Nurse
Male | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Female

Appendix 31 continued



Appendix 32

List of occupation used in Chapter 7, study 2

Female occupations

Receptionist
Nursery Teacher
Nurse
Cabin Crew
Shop Assistant
Legal Secretary

Male occupations

Flight Engineer
Electrician
Labourer
Funeral Director
Car Dealer
Minister/Priest

Neutral occupations

Bar Staff
Psychiatrist
Lawyer
Police Officer
Lifeguard
Real Estate Agent

Appendix 33

Occupation questionnaire used in Chapter 7, study 2

Occupation: Lifeguard

1. **Please indicate the extent to which a lifeguard is a male or female dominated occupation**

Male	1	2	3	4	5	6	7	Female

2. **What would you expect the average age of a lifeguard to be?** _____

Please indicate the extent to which:

3. **A typical lifeguard has broad shoulders, muscular build, rugged appearance, and strong features**

Not at all	1	2	3	4	5	6	7	Very much so

4. **A typical lifeguard has a delicate frame, feminine features, small bones, and soft skin**

Not at all	1	2	3	4	5	6	7	Very much so

5. **Physical strength is required for the job of a lifeguard**

Not at all	1	2	3	4	5	6	7	Very much so

Please indicate the level of each attribute typically found in a lifeguard:

6. **Degree of intelligence of a typical lifeguard**

Not intelligent at all	1	2	3	4	5	6	7	Highly intelligent

7. **Educational attainment of a typical lifeguard**

Low Level of Education	1	2	3	4	5	6	7	High Level of Education

8. **Prestige of the job of a lifeguard**

Not Prestigious	1	2	3	4	5	6	7	Very Prestigious

9. **Trustworthiness of a typical lifeguard**

Untrustworthy	1	2	3	4	5	6	7	Very Trustworthy

10. **Attractiveness of a typical lifeguard**

Unattractive	1	2	3	4	5	6	7	Attractive

The following 6 questions relate to how the typical lifeguard rates on a set of skills, where low indicates a weak ability and high a strong ability.

11. Verbal skills

Low | 1 | 2 | 3 | 4 | 5 | 6 | 7 | High

12. Analytical skills

Low | 1 | 2 | 3 | 4 | 5 | 6 | 7 | High

13. Mechanical skills

Low | 1 | 2 | 3 | 4 | 5 | 6 | 7 | High

14. Spatial skills

Low | 1 | 2 | 3 | 4 | 5 | 6 | 7 | High

15. Clerical skills

Low | 1 | 2 | 3 | 4 | 5 | 6 | 7 | High

16. Communication skills

Low | 1 | 2 | 3 | 4 | 5 | 6 | 7 | High

The following 12 questions relate to the degree to which the job of a lifeguard may require different traits. Please rate how important each trait is for a typical lifeguard.

17. Aggressive

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

18. Sincere

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

19. Independent

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

20. Compassionate

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

21. Efficient

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

22. Sympathetic

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

23. Forceful

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

24. Affectionate

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

25. Reliable

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

26. Has Leadership Abilities

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

27. Happy

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

28. Sensitive to other's needs

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

29. How important is it for a lifeguard to be sociable?

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

30. How important is it for a lifeguard to be emotionally stable?

Unimportant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Important

Appendix 34

Agent rating questionnaire used in Chapter 8

Please rate the following agent on each dimension below. Your ratings should be purely based on physical appearance (rather than in the context of a particular occupation).

Unsympathetic	1	2	3	4	5	6	7	Sympathetic
Cold	1	2	3	4	5	6	7	Affectionate
Insincere	1	2	3	4	5	6	7	Sincere
Untrustworthy	1	2	3	4	5	6	7	Trustworthiness
Unreliable	1	2	3	4	5	6	7	Reliable
Poor Communication Skills	1	2	3	4	5	6	7	Good Communication Skills
Unsociable	1	2	3	4	5	6	7	Sociable
Dependent	1	2	3	4	5	6	7	Independent
Unintelligent	1	2	3	4	5	6	7	Intelligent
Old	1	2	3	4	5	6	7	Young
Inefficient	1	2	3	4	5	6	7	Efficient
Masculine	1	2	3	4	5	6	7	Feminine
Poor Analytic Skills	1	2	3	4	5	6	7	Good Analytic Skills
Weak	1	2	3	4	5	6	7	Strong
Unattractive	1	2	3	4	5	6	7	Attractive
Poor Clerical Skills	1	2	3	4	5	6	7	Good Clerical Skills
Poor Spatial Skills	1	2	3	4	5	6	7	Good Spatial Skills

Appendix 35

Occupation rating questionnaire used in Chapter 8

Please rate how important each attribute is for a <u>shop assistant</u> (where 1 is unimportant and 7 is important)

	UNIMPORTANT						IMPORTANT
Sympathy	1	2	3	4	5	6	7
Affection	1	2	3	4	5	6	7
Sincerity	1	2	3	4	5	6	7
Trustworthiness	1	2	3	4	5	6	7
Reliability	1	2	3	4	5	6	7
Good Communication Skills	1	2	3	4	5	6	7
Sociability	1	2	3	4	5	6	7
Independence	1	2	3	4	5	6	7
Intelligence	1	2	3	4	5	6	7
Age	1	2	3	4	5	6	7
Efficiency	1	2	3	4	5	6	7
Femininity	1	2	3	4	5	6	7
Good Analytic Skills	1	2	3	4	5	6	7
Strength	1	2	3	4	5	6	7
Attractiveness	1	2	3	4	5	6	7
Good Clerical Skills	1	2	3	4	5	6	7
Good Spatial Skills	1	2	3	4	5	6	7

Appendix 36

Matching task used in Chapter 8

Please match each occupation with the agent you think would be **most appropriate in that role**. You may select an agent for more than one occupation.

Shop Assistant **Labourer** **Funeral Director** **Lifeguard** **Minister** **Flight Attendant (Cabin Crew)** **Car Dealer** **Receptionist**

Appendix 37

Kruskal Wallis summary table of effect of attractiveness on preference scores in Chapter 8

Occupation	Kruskal Wallis statistical summary
Cabin Crew	$X^2 = 2.858$, N1 = 4, N2 = 4, p = 0.240
Car Dealer	$X^2 = 1.424$, N1 = 4, N2 = 4, p = 0.491
Funeral Director	$X^2 = 1.086$, N1 = 4, N2 = 4, p = 0.581
Labourer	$X^2 = 0.508$, N1 = 4, N2 = 4, p = 0.776
Lifeguard	$X^2 = 4.960$, N1 = 4, N2 = 4, p = 0.084
Minister	$X^2 = 0.363$, N1 = 4, N2 = 4, p = 0.834
Receptionist	$X^2 = 0.041$, N1 = 4, N2 = 4, p = 0.980
Shop Assistant	$X^2 = 0.910$, N1 = 4, N2 = 4, p = 0.634

Appendix 38

Kruskal Wallis summary table of effect of attractiveness on appropriateness scores in Chapter 8

Occupation	Kruskal Wallis statistical summary
Cabin Crew	$X^2 = 2.710$, N1 = 4, N2 = 4, p = 0.258
Car Dealer	$X^2 = 0.910$, N1 = 4, N2 = 4, p = 0.634
Funeral Director	$X^2 = 0.210$, N1 = 4, N2 = 4, p = 0. 900
Labourer	$X^2 = 0.044$, N1 = 4, N2 = 4, p = 0. 978
Lifeguard	$X^2 = 6.195$, N1 = 4, N2 = 4, p = 0. 045
Minister	$X^2 = 0.363$, N1 = 4, N2 = 4, p = 0. 834
Receptionist	$X^2 = 0.131$, N1 = 4, N2 = 4, p = 0. 936
Shop Assistant	$X^2 = 1.162$, N1 = 4, N2 = 4, p = 0. 559

Appendix 39

Mann-Whitney summary table of male and female participant differences on agent attributes in Chapter 8

	asymp	aaffect	asincere	atrust	areliabl
Mann-Whitney U	164925.5	171546.0	169397.5	170470.5	164669.5
Wilcoxon W	331101.5	351846.0	335573.5	336646.5	344969.5
Z	-1.375	-.219	-.595	-.408	-1.427
Asymp. Sig. (2-tailed)	.169	.827	.552	.683	.154

a. Grouping Variable: gender

acommun	asocial	aindepen	aintelli	aage	aefficie
168603.5	159909.5	157902.5	158632.5	163754.0	166049.0
348903.5	340209.5	338202.5	338932.5	329930.0	346349.0
-.736	-2.262	-2.605	-2.488	-1.583	-1.188
.462	.024	.009	.013	.113	.235

aefficie	aferninin	aanalyti	astrengh	aattract	aclerica	aspatial
166049.0	172586.0	170128.0	148122.5	163169.0	172411.5	167100.5
346349.0	338762.0	336304.0	328422.5	329345.0	352711.5	347400.5
-1.188	-.037	-.474	-4.306	-1.683	-.068	-1.011
.235	.970	.635	.000	.092	.946	.312

Appendix 40

Mann-Whitney summary table of male and female participant differences on occupation attributes in Chapter 8

	osymp	oaffect	osincere	otrust	oreliabl
Mann-Whitney U	71513.500	73593.500	71776.000	67501.000	69863.000
Wilcoxon W	151713.5	153793.5	151976.0	147701.0	150063.0
Z	-1.692	-1.024	-1.618	-3.192	-2.464
Asymp. Sig. (2-tailed)	.091	.306	.106	.001	.014

a. Grouping Variable: c

ocommun	osocial	oindepen	ointelli	oage	oefficie
65997.000	76602.000	63513.500	65184.000	59904.500	71285.000
146197.0	156802.0	143713.5	145384.0	133824.5	151485.0
-3.822	-.065	-4.264	-3.738	-5.431	-1.821
.000	.948	.000	.000	.000	.069

ofeminin	oanalyti	ostrengh	oattract	oclerica	ospatial
59795.500	65045.000	73252.500	57812.000	72673.000	72425.000
133715.5	145245.0	153452.5	131732.0	152873.0	152625.0
-5.715	-3.773	-1.137	-6.212	-1.325	-1.398
.000	.000	.256	.000	.185	.162

Appendix 41

Mann-Whitney summary table of male and female agent differences in Chapter 8

	asymp	aaffect	asincere	atrust	areliabl
Mann-Whitney U	119212.5	125208.0	122616.5	118134.0	128442.5
Wilcoxon W	292378.5	298374.0	295782.5	291300.0	301608.5
Z	-9.366	-8.323	-8.793	-9.582	-7.795
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000

a. Grouping Variable: agentsex

acommun	asocial	aindepen	aintelli	aage	aefficie
122183.0	142263.0	159443.0	142397.0	79905.500	136493.0
295349.0	315429.0	332609.0	315563.0	253071.5	309659.0
-8.886	-5.369	-2.348	-5.350	-16.270	-6.402
.000	.000	.019	.000	.000	.000

afeminin	aanalyti	astrengh	aattract	aclerica	aspatial
9755.500	132284.0	76496.000	107075.5	75132.000	168078.5
182921.5	305450.0	249662.0	280241.5	248298.0	341244.5
-28.480	-7.205	-16.813	-11.496	-17.114	-.850
.000	.000	.000	.000	.000	.395

Appendix 42

Kruskal Wallis summary table of attractiveness differences on agent attributes in Chapter 8

	asymp	aaffect	asincere	atrust	areliabl
Chi-Square	14.568	9.053	32.653	26.428	39.325
df	2	2	2	2	2
Asymp. Sig.	.001	.011	.000	.000	.000

a. Kruskal Wallis Test

acommun	asocial	aindepen	aintelli	aage	aefficie
29.613	17.152	71.785	53.844	14.791	47.852
2	2	2	2	2	2
.000	.000	.000	.000	.001	.000

afeminin	aanalyti	astrengh	aattract	aclerica	aspatial
.255	50.488	83.555	201.225	35.268	34.338
2	2	2	2	2	2
.880	.000	.000	.000	.000	.000

Appendix 43

Mann-Whitney summary table of differences between attractive and unattractive agents; attractive agents and agents of average attractiveness; and unattractive agents and agents of average attractiveness on agent attributes in Chapter 8

Attractive v Unattractive agents

	sympathy	affection	sincere	trust	reliable
Mann-Whitney U	76102.000	71826.000	74713.500	71053.500	60768.500
Wilcoxon W	153130.0	148854.0	151741.5	148081.5	137796.5
Z	-.234	-1.605	-.682	-1.857	-5.176
Asymp. Sig. (2-tailed)	.815	.108	.496	.063	.000

a. Grouping Variable: agent attractiveness

communi cation	social	independent	intelligent	age	efficient	7
60178.500	66631.500	52151.500	56104.500	74587.500	57197.500	
137206.5	143659.5	129179.500	133132.5	151615.5	134225.5	
-5.358	-3.280	-7.913	-6.678	-.719	-6.341	
.000	.001	.000	.000	.472	.000	

feminine	analytic	strength	attractive	clerical	spatial
75260.500	56151.500	50105.000	36407.000	58317.500	59198.500
152288.5	133179.5	127133.0	113435.0	135345.5	136226.5
-.504	-6.722	-8.554	-12.950	-5.956	-5.737
.614	.000	.000	.000	.000	.000

Attractive agents v agents of average attractiveness

	sympathy	affection	sincere	trust	reliable
Mann-Whitney U	66743.500	72409.500	62311.500	60845.500	59264.000
Wilcoxon W	143771.5	149437.5	139339.5	137873.5	136292.0
Z	-3.236	-1.419	-4.667	-5.144	-5.671
Asymp. Sig. (2-tailed)	.001	.156	.000	.000	.000

a. Grouping Variable: agent attractiveness

communi cation	social	independent	intelligent	age	efficient	7
69270.000	75332.500	72136.500	58614.000	63686.500	60074.000	
146298.0	152360.5	149164.500	135642.0	140714.5	137102.0	
-2.444	-.485	-1.515	-5.889	-4.261	-5.438	
.015	.628	.130	.000	.000	.000	

feminine	analytic	strength	attractive	clerical	spatial
75691.500	61401.000	69993.000	68762.000	66880.000	70415.000
152719.5	138429.0	147021.0	145790.0	143908.0	147443.0
-.366	-5.043	-2.197	-2.603	-3.200	-2.091
.714	.000	.028	.009	.001	.036

Unattractive agents v agents of average attractiveness

	sympathy	affection	sincere	trust	reliable
Mann-Whitney U	66331.500	67486.500	60744.000	67115.500	75768.500
Wilcoxon W	143359.5	144514.5	137772.0	144143.5	152796.5
Z	-3.365	-2.997	-5.167	-3.127	-.343
Asymp. Sig. (2-tailed)	.001	.003	.000	.002	.732

communi cation	social	independent	intelligent	age	efficient
67141.500	64937.500	56513.000	72879.500	70269.000	72216.500
144169.5	141965.5	133541.000	149907.5	147297.0	149244.5
-3.113	-3.827	-6.510	-1.274	-2.107	-1.489
.002	.000	.000	.203	.035	.136

feminine	analytic	strength	attractive	clerical	spatial
76795.000	69960.500	55386.000	41720.000	68497.000	65282.500
153823.0	146988.5	132414.0	118748.0	145525.0	142310.5
-.012	-2.244	-6.874	-11.274	-2.682	-3.765
.991	.025	.000	.000	.007	.000

Appendix 44

Means summary table for each agent on agent attribute in Chapter 8

	Sympathy	Affection	Sincerity	Trust	Reliable	Communication
Motorola	4.41	4.19	4.68	4.83	5.12	5.12
Tom	3.20	3.00	3.40	3.29	4.01	4.30
Marc	4.54	4.59	4.44	4.40	4.29	4.79
MP	4.27	4.15	4.67	4.30	3.84	3.48
Charlie	4.45	4.31	4.45	4.44	4.36	4.64
Bruce	3.07	3.26	3.70	3.60	4.18	4.11
Jeff	3.37	4.26	3.02	3.18	3.33	4.78
Ananova	4.08	3.96	4.21	4.20	4.65	5.19
Benoit	3.39	3.59	3.66	3.65	3.81	3.96
Shemail	4.74	5.46	4.96	4.54	3.79	3.97
Cybelle	4.67	4.82	4.71	4.80	4.81	5.45
Tmmy	4.72	4.77	4.71	4.95	4.96	5.48

	Social	Independent	Intelligence	Age	Efficiency	Femininity
Motorola	4.77	5.37	5.65	4.53	5.43	5.82
Tom	3.91	5.12	5.32	4.20	4.83	2.70
Marc	5.53	4.82	4.24	4.64	4.46	2.13
MP	3.83	3.95	3.23	2.05	3.38	1.97
Charlie	4.87	4.73	4.46	5.52	4.49	6.22
Bruce	4.14	5.49	4.98	4.03	4.69	1.77
Jeff	5.66	5.31	3.82	5.38	3.98	1.94
Ananova	5.04	5.16	5.13	5.35	4.95	6.30
Benoit	4.64	4.96	4.18	4.33	4.18	2.21
Shemail	5.29	2.78	3.57	6.76	3.55	6.38
Cybelle	5.32	4.96	4.80	5.57	5.10	5.63
Tmmy	5.50	5.19	4.99	5.79	4.98	6.53

	Analytical	Strength	Attractiveness	Clerical	Spatial
Motorola	5.11	4.05	4.90	5.20	4.42
Tom	4.53	4.29	3.31	3.68	4.50
Marc	3.99	5.29	4.36	3.32	4.17
MP	3.12	4.20	1.85	2.51	3.32
Charlie	4.34	3.46	5.01	4.86	4.03
Bruce	4.33	5.76	4.35	3.28	4.57
Jeff	3.56	5.57	4.63	2.62	4.05
Ananova	4.60	3.70	5.16	4.96	4.39
Benoit	3.80	4.74	3.68	3.06	4.14
Shemail	3.20	2.41	3.89	3.11	3.01
Cybelle	4.54	3.44	4.10	4.93	4.31
Tmmy	4.70	3.33	5.89	5.21	4.23

Appendix 45

Consent form used in Chapter 9

CONSENT FORM

You are invited to participate in a research study conducted by Karen Wilson, from the University of Abertay, Division of Psychology. The purpose of this study is to understand people's perceptions of virtual agents. By considering how people perceive different agents, I can learn more about the best design for an agent, given a particular context.

If you decide to participate, you will be asked to fill in a short agent perception questionnaire, followed by a short interaction session with the agent, via a PC. You will then be asked to fill in 5 follow-up questionnaires. The experimental session will last approximately half an hour.

There are no known risks, discomforts, or inconveniences other than the use of your time known to result from participating in this study. At the end of the study (i.e. after the entire experiment is complete, not after your experimental session today), you will be provided with a complete explanation of this research including the hypothesis. However, I cannot guarantee that you personally will receive any benefits from this research. You can also chose, at the end of the study, to withdraw your data.

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission. Your data will be assigned a subject number that will not be paired up with your name. This consent form will be kept separate from the rest of your data. Data will be made available only to the researcher conducting the study unless you specifically give permission in writing to do otherwise. No reference will be made in verbal or written reports which could link you to the study. If your responses are used as part of a scientific presentation or publication, you will not be identified.

Your participation is voluntary. Your decision whether or not to participate will not affect your relationship with anyone in the Psychology department. If you decide to participate, you are free to withdraw your consent and discontinue participation at any time without penalty.

The experimenter is available to answer any questions you may have regarding your participation. If you have any further questions, please feel free to contact Karen Wilson via e-mail at [redacted].If you have any problem with the experimental procedure or person[redacted]complaint to the experimenter. You will be offered a copy of this form to keep.

Your signature indicates that you have read and understand the information provided above, that you willingly agree to participate, that you may withdraw your consent at any time and discontinue participation without penalty, that you will receive a copy of this form, and that you are not waiving any legal claims, rights or remedies.

****

Print Name: _____

Signature: _____

E-mail address (for debriefing):_____

Date: _____

Appendix 46

Instructions used in Chapter 9

Instructions

In this experiment you will be asked to interact with an interface agent portraying the role of British Airways cabin crew, via text dialogue.

Although the agent is shown on the BA website, the actual webpage is a static image, therefore you cannot navigate around it and use the hyperlinks as though it was a proper website. The only information you can acquire is from the agent.

Your task is to find out times and prices of a flight to New York, as well as information regarding the new visa regulations to the USA.

You can also ask for any additional information about the actual flight (since you are talking to a member of the cabin crew), such as information about in-flight entertainment, meals, or different flight classes.

When you have completed the task, or have been interacting with the agent for more than 10 minutes, the interaction will be stopped and the questionnaires in the folder marked "post" should be completed.

Thank you for your participation ☺

Appendix 47

Agent attribute questionnaire used in Chapter 9

Please rate the agent below on each dimension.

Unsympathetic	1	2	3	4	5	Sympathetic
Untrustworthy	1	2	3	4	5	Trustworthy
Poor Communication Skills	1	2	3	4	5	Good Communication Skills
Unintelligent	1	2	3	4	5	Intelligent
Old	1	2	3	4	5	Young
Masculine	1	2	3	4	5	Feminine
Unattractive	1	2	3	4	5	Attractive
Forgettable	1	2	3	4	5	Memorable
Unappealing	1	2	3	4	5	Appealing
Unfriendly	1	2	3	4	5	Friendly
Useless	1	2	3	4	5	Useful
Dislike	1	2	3	4	5	Like

Overall Impression

Bad	1	2	3	4	5	Good

Appendix 48

System usability score (SUS) questionnaire used in Chapter 9

	Strongly disagree				Strongly agree
I think that I would like to use this system frequently	1	2	3	4	5
I found the system unnecessarily complex	1	2	3	4	5
I thought the system was easy to use	1	2	3	4	5
I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
I found the various functions in this system were well integrated	1	2	3	4	5
I thought there was too much inconsistency in this system	1	2	3	4	5
I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
I found the system very cumbersome (awkward) to use	1	2	3	4	5
I felt very confident using the system	1	2	3	4	5
I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Appendix 49

Communication experience questionnaire used in Chapter 9

	Strongly disagree				Strongly agree
I could readily tell when the computer agent was listening to me	1	2	3	4	5
I was able to take control of the conversation when I wanted to	1	2	3	4	5
It was easy for me to contribute to the conversation	1	2	3	4	5
The conversation seemed highly interactive	1	2	3	4	5
There were frequent and inappropriate interruptions	1	2	3	4	5
This felt like a natural conversation	1	2	3	4	5
I found it easy to keep track of the conversation	1	2	3	4	5
I felt completely absorbed in the conversation	1	2	3	4	5
I had a real sense of personal contact with the computer agent	1	2	3	4	5
I was very aware of the computer agent	1	2	3	4	5
The computer agent was friendly	1	2	3	4	5
The computer agent did NOT take a personal interest in me	1	2	3	4	5

Appendix 49 continued

	Strongly disagree				Strongly agree
I trusted the computer agent	1	2	3	4	5
I enjoyed talking to the computer agent	1	2	3	4	5
I would be interested in meeting the computer agent face-to-face	1	2	3	4	5

Appendix 50

Pleasure questionnaire used in Chapter 9

1. I felt motivated when using this agent

1	2	3	4	5
Strongly disagree				Strongly agree

2. I felt entertained when using this agent

1	2	3	4	5
Strongly disagree				Strongly agree

3. I felt attached to this agent

1	2	3	4	5
Strongly disagree				Strongly agree

4. I felt excited when using this agent

1	2	3	4	5
Strongly disagree				Strongly agree

5. This agent gave me a sense of satisfaction

1	2	3	4	5
Strongly disagree				Strongly agree

6. I could rely on this agent

1	2	3	4	5
Strongly disagree				Strongly agree

7. I had confidence in this agent

1	2	3	4	5
Strongly disagree				Strongly agree

8. I enjoyed using this agent

1	2	3	4	5
Strongly disagree				Strongly agree

9. Using this agent helped me feel relaxed

1	2	3	4	5
Strongly disagree				Strongly agree

10. This agent made me feel enthusiastic

1	2	3	4	5
Strongly disagree				Strongly agree

Appendix 51

Believability questionnaire used in Chapter 9

How believable was the agent?

1	2	3	4	5
Very Unbelievable	Unbelievable	Unsure	Believable	Very Believable

How lifelike was the agent?

1	2	3	4	5
Very Unrealistic	Unrealistic	Unsure	Lifelike	Very Lifelike

Did you believe that the system was of a true technical nature?

Definitely Not	1	2	3	4	5	Definitely

Did you think you were talking to a human being?

Definitely Not	1	2	3	4	5	Definitely

Were you surprised at what the agent understood? What had you expected?

What did you like about the agent? What did you dislike?

Would you use the agent in the future?

During the experiment, did you get the impression that the agent was a real person, or were you constantly aware that you were talking to a computer?

Did the experiment cause you any problems? Was the task too difficult? Did it go on too long? Did you feel pressured or that you were being tested?